OB

14385
.755-K

MENTAL TESTS

MENTAL TESTS

BY

PHILIP BOSWOOD BALLARD

M.A., D.Lit.

AUTHOR OF "OBLIVISCENCE AND REMINISCENCE," AND "HANDWORK AS AN EDUCATIONAL MEDIUM"

HODDER AND STOUGHTON, LTD.

LONDON NEW YORK TORONTO

1920

LB1131
.B3

28567 9
2i

PREFACE

The aim of this little book (its achievement is another matter) is to make the teacher his own critic, and the teacher's critic a just and discriminating judge. By the teacher's critic I mean the head master, the inspector or the examiner—whoever, in fact, passes an authoritative judgment on his work. The teacher and his judge do not always see eye to eye; and the judge, holding as he does the position of authority, is prone to press his point of view. If he cannot convince he commands. And, indeed, if he is to justify his post what else can he do? A difference of opinion about the temperature of the classroom may be composed by an appeal to the thermometer; a difference of opinion about the average height of the pupils may be settled by resort to the foot-rule. But in the realm of mind there are, it is thought, no thermometers and foot-rules: and we sorely need them. We need objective measurements recognised by all as final and unassailable. Indeed, we shall have benefited but little from the new psychology if we are not, critic and teacher alike, made aware of our own complexes—our whims and grooves and fads—

our prejudices masquerading as principles, and our personal standards laying claim to universality. Even if we cannot identify these complexes, to know that they are there is something: it makes us rigorous in judging ourselves, cautious in judging others, grateful for a means of escape from the precariousness of individual judgment. To provide a few such avenues of escape is one of the purposes of the book.

A common malady among teachers, especially women teachers, is over-anxiety about their work. They worry about the progress of their pupils, and are peculiarly sensitive to the opinion of those in authority. And those who worry the most are often those who have the least cause for worry. The best cure for this infirmity is knowledge; knowledge of the pupils, their aptitudes and attainments; knowledge of the best means of measuring progress and assessing results; knowledge of what can be done by other children, and other teachers. And it is hoped that a little wholesome medicine of this kind will be found in these pages.

The solution offered depends ultimately on measurement. There are some who instinctively dislike the idea of bringing measurement into education. They urge that the highest products of education, being spiritual, are outside the realm of time and space to which measurements properly belong. A partial reply to this argument will be found in Chapter I. Here I will merely observe

that it is possible to extend the notion of measurement beyond the physical realm; and that the fact that some of the products of education cannot be measured is no reason why we should measure none. We do, in fact, measure, either well or ill, whenever we examine. If we examine at all we should examine well; and to examine well is to measure accurately.

No claim of novelty is made for the way in which the subject is treated. Much space is devoted to putting before the reader the attempts made in England and abroad to arrive at a scientific system of testing. The author's own modest contributions to the science are to be found mainly in the chapters on Reading and Arithmetic.

In a slightly different form some of the chapters have already appeared in the *Times Educational Supplement* and in the *Journal of Experimental Pedagogy;* and for permission to reprint them I am indebted to the courtesy and kindness of the two editors concerned. The chapter on the Development of Mental Tests was read before the Educational Section of the British Psychological Society, and the chapter on Practical Ability before the Educational Handwork Association.

My personal obligations are many. To the teachers who have so ungrudgingly helped me to standardise the tests; to Professor T. Percy Nunn for his valuable criticism of the chapter on Distribution and Dispersion; to Dr. Emrys Jones for correcting

the proofs—to all these I am deeply grateful. But of all my debts of gratitude by far the biggest is due to my friend Mr. Cyril Burt, who has not only let me use his revision of Binet's Tests, and include his tests of Reasoning and of Spelling, but has carefully read through the whole of the manuscript and given me the benefit of abundant criticisms and suggestions. And all this did he do spontaneously: he volunteered his help when he heard that I was writing on Mental Tests—an act of generosity made all the more striking by the fact that he himself is about to bring out an important book on the same subject.

P. B. Ballard.

Chiswick, 1920.

CONTENTS

CHAP.		PAGE
	PREFACE	v
I.	THE DEVELOPMENT OF MENTAL TESTS	1
II.	INTELLIGENCE AND KNOWLEDGE	22
III.	THE MEASUREMENT OF INTELLIGENCE	29
IV.	BINET'S TESTS OF INTELLIGENCE	48
V.	BURT'S REASONING TESTS	90
VI.	THE MEASUREMENT OF KNOWLEDGE	106
VII.	DISTRIBUTION AND DISPERSION	115
VIII.	READING—	
	(*a*) AS A MECHANICAL ART	134
	(*b*) AS A MEANS OF ACQUIRING IDEAS	145
IX.	SPELLING	155
X.	ARITHMETIC—	
	(*a*) THE FUNDAMENTAL PROCESSES	160
	(*b*) SIMPLE ORAL ARITHMETIC	186
	(*c*) ARITHMETICAL DEVICES	190
	(*d*) APPLIED ARITHMETIC	191
XI.	PRACTICAL ABILITY	196
XII.	COMPOSITION	210
	APPENDIX. SOCRATES ON INTELLIGENCE	221
	INDEX	233

MENTAL TESTS

CHAPTER I

THE DEVELOPMENT OF MENTAL TESTS

It is a truism to say that advance in any branch of physical science is largely dependent on improvement in its mathematics. As the science develops it gets more and more precise. It begins by being qualitative only: it ends by being quantitative as well. Indeed, no generalising idea, however illuminating it may seem, can establish its validity and become recognised as a law of nature, unless it can be proved by the exact measurement of the facts it is designed to unify and explain. Jevons, in his *Principles of Science*, points out that Newton held back his theory of Gravitation for fifteen years because he found that it did not fit the accepted facts about the orbit of the moon. When it was possible to measure the orbit more exactly it was found that the revised figures, instead of refuting his theory, confirmed it to the full. Wisely or foolishly he rejected the undulatory theory of light in favour of the corpuscular because he had no means of measuring with sufficient precision the way in which, and the extent to which, a ray of light is bent out of its course by the edge of an

opaque body. A still more striking instance of the importance of precise measurement is afforded by the recent achievements of Einstein. Everybody, indeed, admits the importance of measurement in physical things, but when it comes to mental things people become sceptical. They admit that measuring would be admirable if it could be done; but it cannot be done. We can measure sticks and stones, but we cannot measure ideas. We can fathom the depth of a well, but we cannot fathom the depth of an emotion. That is the opinion of the frank opponents of mental measurements. The other side is concisely put by Thorndike: "Everything that exists exists in some amount, and if it exists in some amount it can be measured." Here we have the creed of the mental tester: the belief that in some way or other, at some time or other, the most subtle mental processes and the most elusive mental products will be made amenable to measurement. It is not the test itself that is so difficult: it is the evaluating of the product, giving to the result a definite position in a graduated scale, assigning to it, in fact, a number, a mark or a score; and so assigning it that it has objective value. It must not be a guess, but a genuine measurement arrived at by means of a technique which would in everybody's hands yield, within assignable limits of error, the same score.

Mental tests are as old as the human race. A man never consciously and deliberately learns anything without testing himself; he never effectively teaches anything without testing his pupils. All competitive examinations and all school examinations are series of mental tests. And we know that

competitive examinations began in China nearly 4000 years ago, and that school examinations are as old as schools. Then there are the mental tests that figure so largely in legend and story, the riddles that baffle the learning of the learned and are solved by the wisdom of the simple. Œdipus and the Sphinx, Samson and the lion, are types of stories where the test is of mental prowess as distinct from bodily prowess. But these were mere pass-or-fail tests: they were not designed to graduate, but merely to select or sift.

The earliest attempts at real mental measurements were of the nature of indirect measurements. Just as we cannot measure electricity directly, but by measuring something else whose size varies concomitantly with the strength of the electric current can get a result which serves our purpose just as well, so it was thought that some kind of vicarious measurement might be found for the mind. And the most plausible concomitant was the head. This was the idea of Gall, the father of Phrenology. At the close of the eighteenth century he and his disciple, Spurzheim, taught that the head was an index of the brain, and the brain an index of the mind. They did not go so far as to say that a big head meant a big mind, but they did assert that the relative proportions of the skull and the configuration of its surface, would, when exactly measured, yield an exact indication of the mental powers. The palmy days of phrenology were the early decades of the nineteenth century. Nearly everybody then believed in bumps. My readers will recall an account of a certain dinner-party where an unaccustomed guest solemnly asked the company

whether they did not think Milton a great poet; and Charles Lamb, who was present, said he wanted to feel the gentleman's bumps. Indeed, most of us can remember the day when the cultured household was prone to lay great store by a china head mapped out in plots. But we no longer display these things in the home, nor use them in the study. Phrenology has, in fact, failed to establish its claims. If it is science at all it differs from all other sciences by standing still. The charts that we see to-day in O'Dell's window in Ludgate Circus are to all intents and purposes the charts used by Gall and Spurzheim a hundred years ago.

Older even than phrenology is physiognomy. Nearly thirty years before Gall expounded his theory of the skull Lavater put forward his theory of the face. In the year 1772, when Gall was a boy of fourteen, Lavater published his celebrated essay on Physiognomy. His doctrine was that the face, not the cranium, was the index of the mind. A man's fighting qualities, for instance, resided in his nose. A nose that protruded near the root like a Roman nose meant an aggressive disposition. It was the nose of attack. If it protruded in the middle it meant a propensity to fight for others. If it protruded at the point it meant an aptitude for self-defence. The *retroussé* nose was neither inquisitive nor self-assertive, but merely self-defensive.

Shrewd as some of Lavater's observations were, he made the fundamental mistake of confusing the bony structure of the face with its fleshy covering. He regarded the contours of the cheek-bone as much an expression of character as the lines round

the mouth. Many years later, after the science had emerged somewhat from the eclipse into which it had been thrown by phrenology, the study of the face was taken up afresh by scientists like Bell and Darwin. They, however, definitely discarded the osteology of the face in favour of its sarcology, and devoted their attention to the mobile and plastic covering, the changes in which seemed to betoken the prevailing modes of thought and feeling. Physiognomy became, in fact, the science of expression, and only indirectly and suggestively the science of character and intellect. If a man, for instance, has crow's-feet round his eyes before he is forty, it seems no very wild conjecture to assume that he smiles much and is of an amiable disposition. But as weak eyes, as well as a kind heart, may cause crow's-feet, the inference becomes somewhat precarious. Indeed, Darwin never pretended that physiognomy was an exact science. We cannot measure a smile or a sneer or a look of surprise. Heine states in his *Florentine Nights* that he does not know whether the mouths of Parisian women are large or small, because one can never tell where the mouth leaves off and the smile begins.

The next attempt to physiologise the mind was Lombroso's. He was a criminologist who tried to find the bodily peculiarities that went with confirmed evil-doing. He searched for certain marks or malformations which he called stigmata, and which he thought would distinguish the criminal type. A lack of symmetry, for instance, in form or feature was supposed to be one of the common signs of degeneration. We admit its commonness,

for we are all in some degree branded with it. This line of research, however, has fallen into discredit. It is now-a-days believed that if there is a criminal type we must seek its characteristics in the mind and its behaviour, rather than in the body and its anatomy.

The man who gave the greatest impetus to mental measurement in England was Sir Francis Galton. But he was primarily an anthropologist, with a bias towards physical measurements and comparative anatomy. He was imbued with a belief that some sort of correspondence could be found between intelligence and certain bodily traits, such as the length of the middle finger, the character of the finger-prints, and the span of the open arms as compared with stature. And researches in this direction have not proved altogether fruitless. If they did nothing else they gave us the Bertillon system of identification by finger-prints. But they did not bring us appreciably nearer the object of our quest. The physical correlate of intelligence was still undiscovered. It was left to Professor Karl Pearson to deliver the most crushing blow to the belief that we can find a physical scale for mental facts. In 1906 he published the result of an elaborate investigation into the relationship between intelligence and the size and shape of the skull. And his verdict was that the connection between them, if it existed at all, was so slight as to be of no use for purposes of inference. Here ends the first phase of the search for a scale to measure mind. The conclusion reached was definite enough, but entirely negative. We cannot tell a criminal by looking at him ; we cannot tell a genius

by the shape of his skull; and we cannot tell a fool by the length of his ears.

But this does not dispose of physical measurements. For it was next argued that although we cannot measure the mind by measuring the body, we may be able to do so by measuring the powers of the body. A static measurement failed, but a dynamic measurement might succeed. For as soon as the body does something, unless that something is purely automatic, the mind co-operates in the work. Volition, at least, is brought into play. This view led to experiments with the various instruments for measuring muscular strength and skill. The dynamometer, for instance, measures the strength of one's grip, and the ergograph the strength and endurance of the middle finger. But valuable as these instruments were for testing fatigue, they revealed no sort of relationship between mental and muscular traits. The tapping test was more promising. In its simplest form it consists in seeing how many taps per second the subject can make with a pencil on paper. It is found that one can tap faster with the right hand than with the left, and it is a conceivable hypothesis that a very stupid person is, if I may use an Irishism, left-handed all over his body. But although tapping afforded some indication of motor ability it did not signify the presence of any form of intellectual ability.

Reaction-time experiments ran on the same line. It was found that people differed considerably in the rapidity with which they responded to a stimulus; and a delicate apparatus was devised to measure the time that elapsed between the hearing of a sound and the pressing of a button; or between

some other signal and some other motor response. This test is an excellent vocational test. Rapid reaction is important to the boxer, whose aim is to get his fist in first, and to the airman who needs, at the sight of danger, to press the right lever at the right time in response to the right signal.

It will be observed that the quest begins to lose its singleness of purpose. The measurements began by being physical, then they became psycho-physical, ultimately they became almost purely psychical. As soon as the psychical element came in at all, any measurement that was achieved was to a certain extent a psychical measurement. But to measure a particular mental function for its own sake is one thing: to measure it in the hope of finding it an index of general mental ability is quite another thing. The latter has always been the bigger problem; but the former is by no means devoid of purpose and value. To be able to measure any mental function whatever, however limited its operation, is no mean achievement, and in the types of experiments to be now recounted definite measurements have been attained which have proved valuable in developing psychological theory, and serviceable in their application to various pursuits in life. It is only in the larger quest that they have comparatively failed.

In the nineteenth century attempts were made to measure sensations. Weber put forward his celebrated law, which was afterwards interpreted and elaborated by Fechner. It was supposed to establish a relationship between the physical stimulus and the sensation it produced, or, rather, between the way in which an increase in the intensity of the stimulus

was accompanied by an increase in the intensity of the sensation. In the sixties and seventies of last century the Weber-Fechner law gave rise to numerous discussions. While some thought that it revealed the connecting-link between mind and body and was of immense metaphysical import, others held that it was merely an instance of the general law of relativity: it merely illustrated the fact that we judge things not absolutely but relatively. But although the Weber-Fechner law is regarded as of little importance now-a-days—of little importance to the educationist at least—the researches to which it led have considerably advanced the science of mental measurements. Means were devised, for instance, for measuring the acuity of the various forms of sense-perception, particularly of seeing and hearing. More important still, it led to the measurement of sensory discrimination—of the fineness with which we can detect differences in things. For it was suspected that here lurked the clue for which the intelligence-hunters were searching. Indeed, the comparison of lines and the comparison of weights are widely used to-day as tests of intelligence. It was sensory discrimination in the skin, however, that gave rise to the most sanguine hopes. Early in this century an eminent psychologist expressed the belief that he had found a trustworthy instrument for measuring general intelligence. That instrument was the æsthesiometer. The æsthesiometer consists of a pair of sharp points, like the points of a pair of compasses, one of which can be shifted back and fore on a graduated scale. When the points are applied to the skin of a blindfolded subject they are felt as one

point unless they stand a certain distance apart. The least distance apart at which they are felt as two points is the discrimination threshold. It was believed that the sensitivity of the skin, as indicated by this threshold, was a key to the acuteness of the mind; that if a man was thick-skinned he was thick-headed as well. But when the supporters of this theory tried to prove it they found they could not do so. It was found that if sensory discrimination of the skin was an index of anything at all, it was just as likely to be of stupidity as of intelligence; for McDougall and Rivers were able to show that the savages on the shores of the Torres Straits had more discriminative skins than Europeans. So the æsthesiometer was, after all, merely an æsthesiometer and not a phrenometer: it measured sensitivity, but not sensibleness.

The scientific study of memory began with Ebbinghaus, who was the first to measure memory in its simplest form—in the form of retaining and reproducing nonsense syllables—sounds devoid of all sense and all associations. He laid the foundation of modern laboratory methods of testing memory, and of the distinction made in all pedagogical tests between rote memory and substance memory. Galton devised rough-and-ready means of estimating the vividness of visual imagery. Later psychologists have used ink-blots for measuring the rapidity with which images emerge in the mind. A series of blots are shown the subject, and a record made of the time he takes to name the object suggested by each.

It were both tedious and superfluous to take the reader through the various means that have been

invented for measuring the different mental faculties —processes which, in fact, are neither simple nor separate—processes such as attention, perception, apperception, memory, imagination and reasoning. And whenever an instrument or piece of apparatus is used for any of these purposes it is always impossible to claim that it singles out and tests one simple and unitary mental function. One of the most useful pieces of apparatus, for instance, in the psychological laboratory is McDougall's dotting machine. A strip of paper marked with an irregular zig-zag row of small circles passes at a rate regulated by clock-work behind a slot which enables only a small number of circles to be exposed at a time. The subject has to mark with a pencil the centre of each circle as it passes before him. This machine is used to test fatigue, to test attention and to test accuracy of aim; and there may be other kinds of ability that it tests as well.

In the meantime, while specific tests were being rapidly devised and improved, what progress took place in the measurement of general intelligence? This can be best illustrated by reference to the experiments of Mr. Cyril Burt, who has done more to solve this particular problem than any other British psychologist. Mr. Burt began his investigations among school-children in Oxford about the time that Binet published his first series of tests in 1905. His method was to select the group of children in a school who fell within certain age limits (say from twelve and a half to thirteen and a half years old), and to get the head teacher and class teachers to arrange these children in order of intelligence, relying partly on their empirical judgment,

and partly on the results of the school examinations. Then the children were given individually twelve psychological tests, ranging from simple sensory and motor tests to tests of voluntary attention, and it was calculated to what extent the results tallied with the teachers' empirical estimate. The higher the correlation between the two orders, the more satisfactory was the test regarded as an index of intelligence. Some five years later Mr. Burt made a further investigation at Liverpool, using tests of a higher and more complex kind—extending them, in fact, so as to include various types of reasoning. He was thus able to range in a sort of hierarchy, on the basis of their value as criteria of intelligence, a large series of tests involving mental processes of widely varying levels. And the conclusion he arrived at was that when we have arranged them in the order of their complexity we have already roughly arranged them in their order as intelligence-measurers. And of all the tests of intelligence those that measure the power of thinking, that is, the power to understand and to reason, are the best. Thus is common sense vindicated by the psychologist. Thus we have arrived at the conviction that if we wish to test intelligence we must test it directly and not indirectly: we must test those very mental processes which the plain man regards as intelligent. We must note whether the subject is "quick in the uptake," whether he has "nous" or "gumption," whether he can "see beyond his nose." And it will be observed, too, that instruments and machines have been relegated to the psychological laboratory, where they are of inestimable value both to pure psychology and to certain branches of applied

psychology. For educational purposes they are of little use. At any rate, it is certain that no machine has yet been invented which will measure intelligence; and faith in the possibility of such a machine is growing fainter every day. We now regard laboratory tests and school tests as distinct things—a position which was taken up from the very first by that pioneer of educational experiments, Mr. Winch.

At this stage we have abandoned certain specious by-paths which seemed to lead to the desired goal, and we have found the true path, which has, after all, proved to be the common high road which the mass of humanity has been treading all along. But in one sense we are as far from the goal as ever. We know the kind of test to be applied, but we have no scale. We can test, but we cannot measure. It was left for Binet to discover the scale. And the full significance of his discovery is rarely realised. His critics—and they are very numerous—have been so concerned in pointing out what he has not done that they have neglected to give him credit for what he has done. And Binet's crowning glory is, not that he got together a medley of heterogeneous tests for the detection of the feeble-minded, but that he invented a scale. In this he resembles Saul, the Son of Kish, who set out to look for asses and found a kingdom. Binet's scale is his kingdom; not the individual tests—these may so change that there is nothing recognisable left—but the scale itself. Its principle is age-performance. He took a certain test, such as, say, counting backwards from twenty to nought, applied it to a large number of children, found the lowest age at which between 60 and 70 per

cent. of the children passed, and allocated the test to that age. Thus he has a series of about five tests for each age; and if a child of five passes the tests for seven, his mental age is recorded as two years in advance of his chronological age. Thus the unit of his scale was one year of mental age. It was a plan so simple and obvious that one wonders that it never was used before. But it is certain that it never was. We came near it in the old Standards of Examination that were issued by the Board of Education in the days of payment by results. But they just missed it through being an arbitrary scale, based on opinion and not on actual age-performance. And in modern standardised tests the standard is always actually, or ideally, an age-performance.

Binet regarded intelligence as a complex process whose main characteristics were threefold: its purposefulness, its capacity for making adaptation, and its power of self-criticism. He therefore made his tests as heterogeneous as possible. He looked everywhere for tests, and showed much sagacity in finding those that were simple, practical and effective. They involve such ordinary questions as: "Are you a little boy or a little girl?" and such unusual problems as: Put the following words in such order that they make sense—

To Asked Spelling
My I Master
Correct My.

And he used no apparatus. The only material required, beyond what is found in an ordinary school-room, is a series of five pill-boxes similar in size and appearance, but differently loaded.

Binet's scheme led to abundant controversy. Indeed, the literature of the controversy would fill a small library. Everybody abused Binet, and everybody used him. They said his tests were very bad, but they were the best we had. If Binet had lived he would have improved his tests, for he was continually revising them. He, indeed, issued three series—the 1905, the 1908 and the 1911 series—each an improvement on the previous one; and just before he died in 1913 he was engaged in making another revision. His line of research was enthusiastically taken up in America, where three important revisions were issued, the Goddard revision, the Yerkes point-scale revision (where the method of marking was changed), and finally the revision that should be best known in this country, the Stanford revision, carried out by Terman. The basis in each case was Binet, and the alterations and extensions were matter of detail rather than of principle. The tests applied to the recruits for the American Army were partly based on Binet; the tests used for Matriculants at Columbia University were inspired by Binet's, and the tests used in Berlin to select supernormal children are imitative of Binet's. The only extension of principle is the addition of group-tests. Binet's tests are individual, and it takes from half-an-hour to an hour to examine a single subject. A large number of new tests have been devised, which may be applied simultaneously to a large number of pupils; but they agree with Binet's in being age-performance tests, and in being tests of intelligence rather than of acquired knowledge.

The latest development of the Binet system is

to be found in Burt's system of Reasoning Tests; or rather, it is the logical outcome of Burt's own theories, for the only part of Binet's system that he retains is the age-performance basis of selection. The tests themselves he rejects. If intelligence, which Burt defines as "inborn all-round mental efficiency" is mainly manifested in the higher mental processes, the best intelligence tests would embody various forms of reasoning, and would omit all reference or appeal to the lower mental processes. Burt has, accordingly, recently published a series of fifty reasoning tests, which are the best of an initial list of 250 that were tried on miscellaneous groups of children and adults. The questions are not the usual syllogistic questions of traditional logic, but involve the application of thought to the ordinary affairs of life. Where the purpose of the test is to pick out the brightest children rather than to pick out the blockheads, these tests will, in my opinion, prove far more satisfactory than Binet's. They are new as yet and have not been extensively used; but I believe there is a great future for them. They display much ingenuity, and the more difficult tests are far more likely to pick out what Terman calls the superior adult than Terman's own tests.

So much for intelligence tests. But this is only one of the fields in which the technique of mental testing has been improved, and in which it has been put to practical purposes. Laboratory apparatus and laboratory methods have been rapidly improving within recent years, and psychological tests have been applied with great success in the realms of medicine, business and industry. To describe these developments here is outside my scope; I must

limit myself to mental tests in the educational realm, and so far I have not touched the measurement of school attainments.

Our success in securing standardised tests and standardised measurements depends mainly upon three mathematical discoveries. One of these I have already dealt with—that of age-performance. Before Binet's day we tried to measure one unknown by another unknown. We tried to test mental processes by means of tests whose difficulty was merely a matter of guess-work. We knew neither their absolute nor their relative difficulty. Binet gave us a method by which the difficulty of the tests could be graded and standardised. But he himself seems to have made no use of the other two mathematical conceptions which have largely advanced the science of mental measurements—the theory of normal distribution, and the theory of correlation. It was Quetelet, the Belgian mathematician, who showed how widely applicable was the law of normal distribution, what a large number of social and physical phenomena were found to follow the normal curve, and how the form of distribution could be deduced from the laws of probability. But it was Galton who suggested that mental traits generally, and intelligence in particular, would be found to follow the same law. And to-day this curve is universally used to test tests and to certify them. If the results do not conform to the normal curve they at once become suspect. Goddard and Terman have amply demonstrated that Binet's tests triumphantly stand this criterion of validity. To Galton, too, we are mainly indebted for the doctrine of correlation, which has proved so valuable

an instrument in determining the extent to which two orders of measurements tally with each other—or, in other words, the extent to which two functions vary concomitantly. The doctrine of Correlation was developed and elaborated by Professor Karl Pearson, to whom we owe the formula most commonly used. Professor Spearman discovered a formula of correlation which, although not so strictly accurate as Karl Pearson's, is much simpler to use, and is quite satisfactory when the rank only of the subjects is known. This correlation formula was extensively used by Burt, and has figured largely in the exposition and proof of the doctrine of Intelligence as conceived by Spearman and his school.

Another device which has proved of great service is that of "equal grasps," invented by Mr. Winch. Each child in one group is paired with another child of equal merit in the other group. Such an arrangement enables the experimenter to estimate the efficacy of a new method of teaching, or to compare the efficacy of two rival methods; for one group may be taught by one method and the other group by the other method.

It remains for me to deal with the development of mental tests in the realm of school studies—a realm where mental tests in the form of examinations have from time immemorial been regarded as an essential part of the school machinery. But the newer system of school tests differs essentially from examinations. An examination, however carefully it is conducted, is a test of comparative ability. It enables the examiner to arrange his examinees in order of proficiency. But it does not enable

him to compare the proficiency of the group as a whole with the proficiency of any other group as a whole, nor yet with the normal achievement of other pupils under similar conditions of age and schooling. For a public examination is like a rocket : it can be used but once. It measures, it is true ; but it measures the individuals of a group one against the other : it does not measure them by reference to a standard scale. The newer tests, on the other hand, are standardised tests : they are tests whose difficulty is already known in terms of age-achievement. The separate tests, of which an examination consists, are either regarded as carrying equal marks, or as carrying marks which vary in accordance with the examiner's opinion of their difficulty or their importance. Standardised tests, on the other hand, do not depend for the marks they carry on anybody's mere opinion, but rather on their difficulty as experimentally determined. The tests are tested and standardised by a mathematical analysis of the results obtained by applying the tests to a large number of children. The subjective element is reduced to a minimum, and the objective element raised to a maximum.

The extent to which objective tests of school studies are possible depends partly on the study itself, and partly on what aspect of the study it is desired to test. Proficiency in arithmetic, for instance, is easier to test objectively than proficiency in English composition ; and proficiency in the fundamental processes of arithmetic easier than proficiency in the capacity to apply these processes. This difference in the degrees of objectivity is still more readily seen in the attempts that have been

made to measure the various aspects of reading. For proficiency in reading depends upon at least three distinct types of ability—ability to translate symbols into sounds, ability to absorb meaning, and ability to read aloud so as to make the meaning intelligible to others. These are concerned with the mechanical aspect, the intellectual aspect and the elocutionary aspect respectively. The first ability is easy to measure, the second difficult, the third almost impossible. Apart from a few abortive experiments with the dictaphone nobody has ever attempted to invent standardised tests of elocutionary power.

In America alone have educational tests received the attention they deserve. There a number of tests have been devised and norms established in arithmetic, reading, spelling, grammar, geometry and algebra. All these rest on the same basic principle as Binet's Intelligence tests, except that grade-performance is substituted for age-performance. The standardised tests in composition, and handwriting and drawing, however, involve an entirely new principle—a principle which, though not invented by Professor Thorndike, is mainly associated with his name. It consists in devising a scale of typical specimens, a scale arrived at by collating the opinions of a number of independent judges. When it is wished to mark a particular paper, it is compared with the specimens of the standard scale with a view to discovering to which it is most nearly equal in merit. If the variability in the marking of the same set of papers by a number of independent teachers is less when the scale is used than when it is not used, it is claimed that

the use of the scale is justified. As a matter of fact it is found that after some practice in the use of the scale the variability *is* less; but not so much less as to give this type of measurement any very high degree of objective validity. Even in America the scales are not used much. In England they are not used at all. But the other scales are to a certain extent. We have not, in this country, gone far in the invention of tests and the establishments of norms; but we have begun the work, and we are slowly but steadily forging ahead.

CHAPTER II

INTELLIGENCE AND KNOWLEDGE

All who have given serious thought to education have been wont to exalt intelligence at the expense of knowledge. "Wisdom," says an ancient writer, "is the principal thing; therefore get wisdom; and with all thy getting get understanding." And although by wisdom he doubtless meant something more than intelligence, he at least meant intelligence. On the other hand, "he that increaseth knowledge increaseth sorrow." Here in these far-off times we see recognised a distinction which has much exercised the mind of the modern educationist—the distinction between wisdom and learning, intelligence and knowledge, mental power and mental content.

The cry for intelligence was acutely clamant in the first decade of the twentieth century. The primary schools had just recovered from the cramping effect of annual examinations and had begun to breathe more freely. Intelligence became a cult and a quest and a watchword. Teachers aimed at it; inspectors looked for it; administrators encouraged it. There was a "slump" in accuracy and a "boom" in intelligence. And the educational psychologist, seeing herein a promising field for his labours, began to investigate intelligence and to

cast about for means to measure it. Although Binet's name was alone prominent before the public, he was by no means the only investigator; nor, perhaps, the most original. Still, the work he achieved was great. He had a genius for discerning what was worthy in the work of others; he cast his net so wide that nothing valuable escaped him; and—most important point of all—he kept a steady eye on practicability.

But while the teacher tried to cultivate intelligence, and the psychologist tried to measure intelligence, nobody seemed to know precisely what intelligence was. It was certain that the term covered a wide field. When at a recent meeting of an education committee a member asserted that teachers as a body were highly trained and intelligent, an opponent (*advocatus diaboli*) retorted that the same was true of elephants. Clearly there was need to sharpen and define the term. It is not enough to say that it means ability as distinct from knowledge, capacity as distinct from content, power as distinct from product. For how can we gauge a mind's capacity except by finding out what it can contain? But what it contains is knowledge. And how can we measure a mind's power without measuring its product? But its product is, again, in part at least, knowledge. Binet saw this difficulty, and frankly accepted knowledge as one of the many marks of intelligence. One of his earlier intelligence tests, for instance, for a child of nine is: name in order the days of the week. He would probably argue that a child of nine who in his intercourse with his fellow-creatures had not picked up these names was necessarily somewhat stupid,

Binet was inclined to adopt the easy-going attitude of the artist, who does not mind a notion being nebulous so long as it is workable, rather than that of the scientist, who strives for precision and consistency at all costs.

There are other psychologists, however, who have taken up the question in a different spirit; most notably Professor Spearman in England and Professor Thorndike in America. Their mode of research was mathematical. They submitted specific mental abilities—the ability to add numbers, to memorise words, to discriminate lengths, to sort cards, and so forth—to certain rigid tests, and carefully marked the results. Then they compared the various scorings and found the correlation between them; that is, the degree of concomitant variation, or the extent to which the compared abilities tend to rise and fall together. And from similar statistics the two investigators arrived at entirely different conclusions. Thorndike concluded that the individual abilities were entirely independent; that there was, in fact, no such thing as general intelligence, but only particular intelligences. For may not a man be intelligent at geometry and stupid at history, or brilliant as a poet and hopelessly bad at figures? Spearman, on the other hand, concluded that there was a certain dominant factor common to all the specific abilities, a central fund of intellective energy, to which the term general intelligence or general ability might fitly be applied. To put it in another way, Thorndike held that a man's intellectual wealth consisted entirely of coupons; Spearman that it consisted partly of coupons, each of which may be expended in one

direction only, and partly of cash, which could be expended in any direction. Both views recognised the obvious fact that individuals varied widely in intellectual wealth; the views clashed on the question of the availability of that wealth.

Let us examine the ground of the belief in general intelligence. It is a matter of common observation that a man who is good at one thing is good at most other things. At least it is so as a rule; cases of one-sided ability are very rare. Generally speaking, a wise man is wise in all things, a fool is a fool all round. Indeed, it can be proved mathematically that there is a positive correlation between all forms of native ability; they always tend to hang together; the odds are always in favour of high ability in any given function being accompanied by high ability in any other function. Why should this be? Why should mathematical ability be positively correlated, as it is, with linguistic ability? Even if we make every allowance for such operations as might be conceived to be common to the two abilities, we still fail to account for the whole relationship. There still remains an unexplained nexus. We are forced, in fact, to assume a general factor common to all the multifarious operations of the mind, a factor with which each specific ability is, in its own measure, charged and energised. This common factor is intelligence. Such is Spearman's reasoning; and its cogency is hard to dispute. But he further arrives at the disconcerting conclusion that this central factor cannot be cultivated. It is born with one, and can neither be improved by schooling nor dulled by neglect. Intelligence is mother wit, and mother wit is a

matter of heredity. The ancient writer already quoted holds a more hopeful view. He believes in the possibility of cultivating wisdom. But even he admits that it is sometimes difficult : " Though thou shouldest bray a fool in a mortar among wheat, with a pestle, yet will not his foolishness depart from him."

It follows from this that the only practical thing a psychologist can do with general intelligence, if there is such a thing, is to measure it. Hence the assiduous pursuit of mental tests. It is clear that general intelligence cannot be directly tested, for it can never be found alone : it is always embedded, as it were, in specific abilities. But can it not be tested indirectly ? We have in the thermometer an excellent example of indirect measurement. There are no known means of measuring temperature directly ; we have to make use of the fact that the expansion of a thin column of mercury is almost perfectly correlated with temperature. So instead of measuring the temperature, we measure the mercury ; which does quite as well. Now is there not some simple function of mind or body which when measured will give us in the same way an exact valuation of intelligence ? The quest of this key-ability resembled the quest of the philosopher's stone. It sometimes led to the discovery of unexpected treasure ; it often led to the discovery of mare's nests ; it always, as far as the essential search was concerned, ended in disappointment. Every device failed that assumed an essential correspondence of soul with sense and sought to measure mind through matter.

There was no help for it. The hope of finding

a simple key-ability had to be abandoned; and investigators had to fall back on the laborious expedient of testing as many abilities as possible and, by a process of mathematical analysis, extracting the common intellective element. Among the pioneers in this particular field of research, Mr. Cyril Burt takes the foremost place. After years of patient labour he proved that while almost any kind of ability was a presumptive sign of intelligence, some abilities were much safer signs than others. To use his own words: "Of all the tests proposed, those involving higher mental processes, such as reasoning, vary most closely with intelligence."

There are no tests of intelligence that are more widely used than Binet's. And Binet's tests are based on the principles I have just expounded. Like Spearman and Burt, he discarded brass instruments, and relied more on the higher mental processes than on the lower. His tests measure general ability simply because they measure a large number of specific abilities. So far as he is concerned, it does not in the least matter whether Spearman is right or Thorndike; on either theory his tests are a real, though rough, measure of individual mental endowment.

William James, in advocating his pragmatic method, contends that the best way to discover the essential difference between two conflicting theories is to find out the practical difference in the consequences that flow from them. How, in fact, do they affect practice? We have seen that, whichever of the rival theories of Spearman and Thorndike be accepted, it makes no difference in the mode of testing; does it make any difference to the teacher?

Again the answer is "No." On either view the only cultivable thing is the multifarious group of special abilities. It must not, however, be thought that all this theorising and researching has had no effect on school practice. It most distinctly has. Its effect has been to broaden the outlook, to multiply the school pursuits, to vary and amplify the methods of study.

CHAPTER III

THE MEASUREMENT OF INTELLIGENCE

The British Press refers to mental tests as though they were new things invented by Americans. In point of fact they are neither new nor American. They have been the common property of the race since the dawn of history. They are no more mysterious than the conundrums that delight children at a Christmas party. They are, in fact, merely questions or tasks that invite a trial of intelligence. What is new is not the tests themselves, but the aptness with which they are chosen and the scientific precision with which they are applied.

A tacit distinction seems to be made between examinations and mental tests. This distinction is illegitimate; for an examination is nothing but a series of tests, which are just as mental or psychological as any that have ever been devised. They are not non-mental tests, but simply a special kind of mental test. The real distinction lies between tests of knowledge and tests of ability; tests of school attainments and tests of natural intelligence; tests of book-learning and tests of mother-wit—a distinction which is easy to make but difficult to maintain. For it is impossible to devise a test of ability which does not also test knowledge, and

impossible to devise a test of knowledge which does not also test ability. Let us, for example, take one of the most characteristic of Binet's intelligence tests. The child is asked to say what is absurd in the following: "If I should ever grow desperate and kill myself, I will not choose Friday, because Friday is an unlucky day and will bring me unhappiness." To answer this correctly he must know, among other things, the meanings of the words, and he must know that a dead man is, in this world at least, neither happy nor unhappy. If the question were put in these words to a unilingual Chinaman he could never answer it, however intelligent he might be; and to a race of immortals it would be meaningless. Now let us consider this pedagogical test: Which are the four largest towns in Scotland? A correct answer would at least involve ability to grasp a fact and to remember it, to understand a question and to respond to it. The difference between the two types is one of proportion. To answer the second we must acquire a special bit of knowledge which casual experience will not necessarily force upon us; to answer the first involves the application of such knowledge as no sane mortal can fail to pick up in the ordinary course of life.

The distinction, however, so far as it goes, is perfectly sound, and indeed is constantly cropping up in the folk-lore and legends of all races. The point of the old English ballad of King John and the Abbot depends entirely on this distinction. The king sets the abbot three mental tests. At the risk of forfeiting his life in case of failure, the abbot has to say what the king is worth, how long it

would take him to ride round the world, and what he is thinking of. A respite of three weeks is given. The abbot cudgels his poor brains in vain. He rides to "Cambridge and Oxenford, but never a doctor there was so wise that could with his learning an answer devise." In fact, the only man who could rescue him from his plight was his own shepherd, a man who could neither read nor write. The shepherd, who happened to resemble his master in appearance, disguised himself as the abbot, presented himself before the king, and spoke thus: "How much are you worth? Twenty-nine pieces of silver, for you are worth at least one piece less than our Saviour. How long will it take you to ride round the world? You must rise with the sun, and ride with the sun, until he rises here again next day; then the journey will take you exactly twenty-four hours. What do you think? You think I'm the abbot; but I'm not: I'm his shepherd." Here we have three tests, which all the learning in the land could not satisfy, triumphantly solved by simple mother-wit. And, indeed, all through the records of folk-lore and mythology do we find the learning of the learned put to confusion by the wisdom of the simple. The supreme ordeal which reveals the native nobility of the hero is just as likely to be the solution of an enigma or the interpretation of a dream as the slaying of a dragon or a giant. While the dragon is the test of physical prowess, the enigma is the test of intellectual prowess. Perseus by his courage rescued Andromeda, but Œdipus by his intelligence rescued the whole of Thebes. As the reader will remember, the riddle of the Sphinx ("What animal walks in

the morning on four legs, at noon on two, and in the evening on three?") was solved by Œdipus alone, a king's son reared by peasants, a man in whom nature triumphed over nurture.

The whole family of riddles, puzzles, charades and conundrums belongs to a larger group of ability tests. They require for their solution no special knowledge, but a general aptitude for applying knowledge—for the discernment of subtle analogies and contrasts. We must not, however, fall into the error of thinking that the modern application of mental tests consists in the mere asking of riddles. The tests are, in a sense, riddles, but they are riddles of a special kind: they must have diagnostic value. Let me illustrate by comparing an old riddle with a new test. The riddle is: What relation does a loaf of bread bear to a steam-engine? And the test: Complete the analogy: As a loaf of bread is to a glass of water, so is eating to ——. The answer to the riddle is "mother"; for a loaf of bread is a necessity, a steam-engine an invention, and necessity is the mother of invention. This is not a test of intelligence, but of a perverted ingenuity. It is not intended to gauge one's reasoning powers, but to raise a laugh. As a means of diagnosis it is quite useless, unless, indeed, one gives the result a negative interpretation; for success in its solution is, as a mark of sanity, no more significant than failure. It is sophistry rather than logic, a jest rather than a test. Very different is the analogies test given above. Failure to answer this would, in an adult, mean a serious defect of intellect. For its solution lies not in some crooked by-path of sophistry, but on the great highway of human

thought. That is why it serves for a diagnosis of mental endowment and unfoldment. For there is a certain stage of mental development below which it cannot be answered, and above which it cannot be mis-answered. Although, therefore, the enigmas of the ancients and the catch-questions of to-day are, in the broad sense of the term, mental tests, they lack the special characteristics of those tests which are now-a-days called mental or psychological. The whole point and pungency of the riddle or the catch lies in its being exceptional. To repeat it in another form is to thwart its purpose. For its purpose is to trip us up; and once the trick of it is known it ceases to trip us up. After we have got the hang of such a sentence as "Pas de l'yeux Rhone que nous," it is quite easy to read "Guy n'a beau dit qui sabot dit nid a beau dit-elle?" When Œdipus, in pondering over the riddle of the Sphinx, guessed that "the morning" meant "the morning of life," the enigma was virtually solved; and all others based on the same metaphor. The riddle, in fact, needs novelty for its success. The nut to be cracked may be any nut but a chestnut.

As distinct from this spurious mental test, the genuine mental test depends for its value upon its universality. It is always a member of a family, and the larger the family the better. The analogies test given above may be multiplied indefinitely; indeed, as many as a hundred have been set at one sitting. For example: *As* cat *is to* kitten, *so is* dog *to* ——? *As* sheep *is to* ox, *so is* flock *to* ——? *As* Paris *is to* France, *so is* London *to* ——? In testing ability, as in testing knowledge, it is always best to set a fair number of questions. To base a

judgment on one test, however good that test may be, is extremely precarious; when the test itself is suspect, no verdict at all can be reached. Each test must itself be tested, and the results interpreted in accordance with the laws of probability. I have, for instance, tried the following test on young and old: If you are given a corked bottle half full of wine, how can you get the wine out without taking out the cork or breaking either the cork or the bottle? Most adults proclaim the task impossible, but many a small boy, who has solved certain difficulties with his bottle of ginger-beer or liquorice water, can answer the question promptly. The test, in fact, proves to be too dependent on an accident of experience to be a real gauge of practical ability. Nor is it necessary to fall back on such doubtful tests when we have so many that are of demonstrable validity.

Another element of prime importance in scientific testing is speed—an element which is clearly exemplified in the cancellation test. The subject is required to cancel as rapidly as possible all the *a*'s or all the *r*'s, or any other letters, separately or concurrently, in a page of printed matter. A limited time is allowed—say, two minutes. In rating the results the marker has to take into account the number of letters rightly erased, the number wrongly erased, and the omissions. It is obvious that with unlimited time all who could distinguish the letters at all would score full marks. To ignore the time element is to nullify the test.

If, however, there is one single feature which essentially distinguishes the modern test from the ancient, it is that the modern test is standardised.

The man who first convinced the world of the necessity for standardised tests, as distinct from casual and haphazard tests, was Alfred Binet. And he achieved it incidentally. What he really was after was to find out which children in the schools of Paris were mentally defective. To do this he devised two carefully graduated scales, one to measure school attainments and the other to measure general intelligence. The first measured roughly, the second more exactly; the first sifted out the suspects, the second found from among the suspects the real blockheads. The assumption was that all fools are dunces, but all dunces are not fools—an assumption quite in accordance with common sense. This general scheme, full of flaws as it is in detail, has given birth to the whole modern system of scientific testing. He has not only shown us how to discover intellectual and moral weaklings, but he has led us into a fertile land which it is our duty to explore and subdue.

Binet avoided the mistake of trying to deduce a scale from first principles. He based his scale on fact. Before asserting what degree of intelligence a child of ten ought to possess, he took the trouble to ascertain what degree of intelligence a child of ten actually does possess. Before testing children with a test, he first tested the test with children. By applying it to a large number of children of different ages he was able to fix the lowest age at which the majority were able to pass it. Let me illustrate this with reference to a type of test which was evidently a favourite of Binet's, for in his scale of fifty-four intelligence tests it appears four times —more frequently, in fact, than any other. It

consists in the repetition of a series of digits. The examiner says: "I am going to say x numbers. Listen and repeat them after me: 5, 8, 2, . . . etc." The numbers are uttered steadily at intervals of half a second. Binet found that at three years of age few could repeat three digits, but considerably more than half could repeat two. Thus he fixed the repetition of two digits as a three-year-old test; and acting on the same principle he assigned three, five and seven digits to the ages of four, eight and fifteen respectively. This innocent little test is not, in its implications, so simple as it seems. It measures the span of primary memory—the number of things one can hold in the conscious memory at the same time. It depends on the fact that consciousness is not a point but a patch; and the assumption is, the bigger the patch the bigger the mind. When the mind ceases to attend to A and passes on to B, A does not suddenly vanish from consciousness leaving B in sole possession, but fades gradually away, forming in the process what William James calls a "fringe" to B. If by the time the last digit is uttered by the examiner the first has quite disappeared from the child's mind, the child will fail in the test. His psychic fringe is too small.

The importance of primary memory is more apparent in dealing with words than in dealing with figures. For if all the essential parts of a spoken sentence do not in some sense reverberate together in the mind, the sentence has no chance of being understood. Fortunately the words in connected discourse have greater cohesion than a string of digits, and the number of syllables that

can, according to Binet, be repeated at the ages of three, five and fifteen amount to six, ten and twenty-six respectively. Roughly speaking, a child should be able to repeat twice as many syllables as the number of years he has lived. Binet's typical example for fifteen is: "The other day I saw in the street a pretty yellow dog. Little Maurice has stained his nice new apron."

Much interest is attached to Binet's third test for children of twelve years of age: "I am going to allow you three minutes, and I want you to say as many words as you can think of. Some children have said more than two hundred; let me hear how many you can say. Ready? Start." In order to pass the child must say at least sixty words. This is an ingenious way of finding the child's association time—the time it takes one word to call up another. The laboratory method is to deal separately with each word and to measure precisely by means of a specially constructed clock the interval that elapses between the hearing of the word and the emergence of an associated idea. And, indeed, some such plan is necessary in scientific investigations of the association process itself. But for measuring mental activity in children, for finding roughly the rate at which ideas march through the mind, Binet's method is equally effective and much simpler. Although he uses this test once only, it has been found that the number of words a child can utter in the given time increases steadily with his age; that norms or averages could consequently be fixed, and the test used to mark different levels of intelligence. It has also been found that one minute will give results just as trustworthy as three

minutes. All this serves to illustrate how Binet's tests are constantly being criticised and refined upon; how tests themselves are not exempt from the testing of time and experience.

Interesting, however, as these tests are, they are probably inferior in diagnostic value to the absurdity tests to which I have already referred. And although Binet sets them at one point only—ten years of age—they are appropriate to all ages except the very lowest. The following test, which I devised some years ago to arrange rapidly (and of course roughly) in order of intelligence a group of children of eleven years or over, was found to be suitable for all children who could read fluently and were advanced enough to express their ideas in writing—

John Carew lived in a small cottage which stood on the top of a barren hill and faced the east. From the foot of the hill a grassy plain stretched in every direction as far as the eye could see. On the evening of John's thirtieth birthday, while he was sitting on the front door-step looking towards the setting sun and watching his shortening shadow on the gravel path, he suddenly became aware that a horseman was riding down to the cottage. The intervening trees and foliage made it difficult for him to see clearly, but he was able to perceive that the horseman had only one arm. When, however, he got a closer view he recognised that the visitor was his son William who had left home to join the army twenty years before and had not been heard of since. On seeing his father William immediately dismounted, ran towards him, threw his arms round his neck, and burst into tears.

Each child in the class was given this passage in print, and was allowed a quarter of an hour to read it and to write out any absurdities he could find therein. It contains about seven absurdities, and it was found that it did actually differentiate the children in close accordance with the teacher's rating of their general intelligence. Elementary school children of eleven years of age who discovered at least five absurdities were, as a rule, those who on other grounds had been selected as worthy of a secondary school training.

It has been complained that most absurdity tests are too easy for the older pupils. Given, however, a time limit, it would not be difficult to devise a test which would floor the majority of adults. Let the reader, for instance, try to find within fifteen seconds the absurdity (there is but one) in the following: "John Jones, who had married his widow's sister, used to say that if a man had a bad sister it was his misfortune, but if he had a bad wife it was his fault."

When Binet published his scale in 1908 he regarded it as tentative. Indeed, he revised it himself three years later; and since his death many a patient experimenter has laboured at separating still further the gold from the dross. It has been found that while some of the tests are almost worthless, others are of no small value: they are reliable in practice and suggestive in theory. Moreover, new tests are daily being devised and put to the trial of rigid experiment, so that we are gradually accumulating a body of tests which bear the hall-mark of demonstrated success.

One of the main defects of Binet's scale is its

bias in favour of language—of words rather than things; another defect is the paucity of tests (there are but five) prescribed for adults. When therefore adults have to be tested for intelligence—and especially such illiterate adults as were plentiful in the American Army—recourse has to be had to tests of a new type. One of the most useful of these is the "construction board," a variation of the jig-saw puzzle. The sections, however, are generally rectangular, and the time always restricted.

A test which has gained considerable vogue during recent years consists merely in the giving of instructions to be obeyed by the subject. These instructions may possess any degree of complexity and may be used to mark the rapidity with which the meaning is grasped. For instance: "Print your Christian name (or names) in small letters and your surname in capitals unless there are more than six letters in your surname, in which case you should print your surname in small letters and your Christian name (or names) in capitals." It is quite easy to see how these tests may be indefinitely complicated either by making the provisos more puzzling or by simply increasing their number. As a sample: "If your grandfather's only child was your uncle draw a square; if not, draw a circle." Here, as always, a time limit is essential.

So much for intelligence tests. Binet's other scale (his *barème d'instruction*) consisted of ordinary examination questions, with the difference—a highly important difference—that they have been carefully sifted and standardised. The development of this scale is quite another story. Nor must the intelligence tests with which we have been dealing be

confused with vocational tests, which are designed to find out what sort of work a person is best fitted for. These are just as often physical tests as mental tests, and frequently involve the use of delicate and complicated apparatus.

Having made a distinction between the two broad types of mental tests (tests of knowledge and tests of ability), and having shown that these types are never quite pure, let us examine the bearing of this theoretical distinction on the practical art of examining. There are four possible methods of procedure. First, we may dispense with the ordinary examination altogether and substitute a series of ability tests. This is no new proposal. It has been made many a time before, but it has always been rejected. And rightly so. For it cannot be too strongly insisted on that education is directly concerned not with natural ability but with culture. For natural ability, as scientifically conceived, grows with the growth of the brain, fluctuates with freshness and fatigue, and varies with varying states of health and nutrition. The most that the schoolmaster can do is to take full advantage of what natural ability his pupils may happen to possess. His business is not to train intelligence but to use it—to use it himself, and to see that his pupils use it. And his success is measured by the degree of culture he imparts. But culture means knowledge. True, it means other things as well; but it means knowledge at least. And the most effective way of testing knowledge is by examinations.

The second possibility is to leave things as they are—to rely on examinations pure and simple, examinations of the good old-fashioned sort. But

this policy implies a wilful blindness to certain grave defects which have been pointed out repeatedly and persistently for the last half-century.

The third possibility is to hold two distinct examinations, one a pedagogical examination, and the other a psychological examination—one a test of acquired knowledge, the other a test of natural aptitude. This is virtually Binet's plan for detecting the feeble-minded. The teacher sifts first, the psychologist afterwards. It is also the scheme that is reported to have been recently adopted at Columbia University for selecting candidates for entrance. Those who pass the ordinary matriculation examination are further submitted to a series of tests for general intelligence. It is not quite clear whether any are knocked out in the second round. It is not indeed clear whether the second round is intended to knock them out at all. The probability is that the second examination is not selective but experimental. It is intended to reveal the degree of correspondence between the results of the two types of examination. Rash indeed would be the examiner who would lightly brush aside the evidence of combined scholarship and intelligence which a wisely conducted examination affords. Besides, the virtue of matriculation lies in the fact that it guarantees a certain minimum of culture. The university starts no course of study from the beginning; it takes up the tale at the point where the secondary school left off. It demands of its *alumni* a determinate measure of literacy, and by means of a matriculation examination it assures itself that its demand is met. And if this demand is met—if the candidate has sufficient intelligence

to carry him successfully to that point in the road of learning—it would need more proof than our modern mental tests can at present afford to demonstrate that he can go no farther. If the purpose of the second examination is to reduce the element of chance, there seems to be more reason for re-testing the failures than for re-testing the passes. For failures may be due to illness or excessive nervousness, or sheer bad luck. Moreover, a high degree of intelligence in a candidate may well be regarded as counterbalancing certain gaps in his store of knowledge.

The last alternative is to combine both types of test in one examination. When this took place in the past it was mainly a matter of accident; now it is done by design. It is the way in which the new science of testing has affected English examinations. The examiner is beginning to ask himself what it is that he is really examining. Is it parrot knowledge, or is it knowledge that has been intelligently acquired and can be intelligently applied? Is it the application of common knowledge in an unfamiliar field, or is it the application of exotic knowledge to familiar instances? Is it the capacity to acquire, or the capacity to express? In any case the wise examiner never neglects to test the capacity of the candidate to apply knowledge and to express it in new forms, for in so doing he kills two birds with one stone: he tests both knowledge and intelligence. Speaking generally, the younger the pupil the less the importance to be attached to school attainments as distinct from native ability. In other words, if intelligence tests are important anywhere, they are important at that

stage in the pupil's career when we decide upon the type of education for which he is best fitted. This principle is clearly observed in the selection of mentally defective children at the age of seven; less obviously in the selection of supernormal children at the age of eleven. It is customary at examinations for junior scholarships to set two papers only, one in English and the other in arithmetic. And the questions are so devised as to frustrate, as far as may be, the designs of the crammer. That this examination will radically change in the near future is highly improbable, for it is difficult at present to conceive a better practical device for testing annually the intelligence of a large number of children. The fact that it is conducted every year, or even twice a year, in the same schools renders known and standardised tests almost useless. If we are to remove entirely the possibility of special coaching, the element of novelty is essential. And the scope for variety in the realm of English and mathematics is virtually unbounded. Here intelligent anticipation on the part of parents and teachers can readily be defeated. Again, the 8000 children who sit every six months for the London junior scholarship examination can all be examined quite easily in one morning with no superintendents besides the ordinary teachers. Under the Binet scheme, where the testing is oral and individual, it would take a trained examiner nearly four years to get through one half-yearly batch. To test the whole lot in one morning would need the services of 1400 examiners specially trained for the purpose. As a matter of fact samples of the candidates have been tested by both written

and oral methods, and the results obtained are so similar that confidence in the present system is amply justified.

There are two fairly distinct types of ability which are tested by these junior examinations—mathematical ability and literary ability. Proficiency in problem arithmetic is in itself no poor criterion. One of the most trustworthy of Binet's tests is the second in his scale for eight-year-olds: Count backwards from twenty to nothing in twenty seconds. It is the departure from the beaten track that gives it its value. The modern tendency, however, is to attach more importance to the English paper than to the arithmetic. Binet long ago remarked that no child who could compose was mentally defective; and Mr. Cyril Burt, the psychologist to the London County Council, has expressed the opinion that the school subject most clearly symptomatic of intelligence is composition, provided it is marked for its power of thought. The only serious defect of the present system is the absence of tests of the third broad type of ability—manual ability. This defect has been carefully considered by the London authority, who have adopted the view that in early years the ablest with their heads are also as a general rule the ablest with their hands; that it is very difficult to discover at the age of eleven (that is, before the children have begun to attend the handicraft and domestic centres) what practical ability children actually do possess; and, finally, that those few children who have a special aptitude for craftsmanship and escape the junior county scholarship net are captured by the trade scholarships.

It were unwise to prophesy what developments

will take place in English examinations in the near future. Already one English authority has included in its junior examination a test of the "instructions" type given above, and it is not impossible that a third paper will become the general rule—a paper corresponding to the obsolete general knowledge paper, except that it will aim at discovering not whether the candidate is well-informed, but whether he is sharp-witted. The secondary schools and universities show as yet no alarming symptoms of infection.

Thoughtful people have recently been asking: Why is it that America has been moving so rapidly in the matter of mental tests while England has almost stood still? The answer is simple: Speaking generally, Americans believe in psychology but Englishmen do not. When America entered the war one of the first things she did was to mobilise her psychologists. The war was nearly over before England discovered that psychologists were of any use. And the discovery was due partly to the witnessing of what was achieved in the American Army, and partly to an appreciation of the wonderful results that followed the psychological treatment of shell-shock and war-strain by those numerous psychologists (medical and non-medical) who volunteered their services early in the war. The fault did not lie with the British psychologists. In the pursuit of their particular science and in the invention of new tests they were in no way behind their neighbours. And the Americans knew this; they consulted our professors and freely used their tests. There is no psychological instrument which is more generally useful than the dotting machine, an

invention of Mr. McDougall, of Oxford; there is no instrument more specifically useful than the K tube, an invention of Dr. C. S. Myers, of Cambridge. The new science of testing could never have reached its present stage but for the discovery of that potent means of mathematical analysis known as "correlation." It was discovered by an Englishman, perfected by an Englishman, and simplified for common use by an Englishman. In the investigation of general intelligence and in the critical examination of the methods used nobody has done better work than Mr. Cyril Burt. Like Binet, he was concerned in testing tests as well as in testing ability. But the methods he adopted were more rigidly mathematical. He invented tests too; among others the analogies test mentioned above. Many other Englishmen have successfully laboured in the same field, such as Mr. W. H. Winch, Dr. W. Brown, Dr. E. O. Lewis, Professor J. A. Green, Dr. J. L. McIntyre, Dr. N. Carey, Miss N. Taylor, Mr. H. B. English, Miss May Smith, Dr. Bernard Hart, and Dr. Edgar Schuster and other investigators in the psychological laboratories of University College, King's College and Oxford. Such records of their work as have already been published may be found distributed among the pages of the *British Journal of Psychology*, the *Journal of Experimental Pedagogy*, and the Annual Reports of the British Association.

CHAPTER IV

BINET'S TESTS OF INTELLIGENCE

The translation of Binet's Tests given below was made by Mr. Cyril Burt in consultation with Dr. Simon, who was Binet's collaborator in devising and standardising the scale. The tests, originally standardised for Parisian children, have been tried by Mr. Burt on a large number of London children, and modified to suit their characteristic rate of development. The tests, in fact, are here rearranged in the order of their freshly ascertained difficulty, and the age-assignments are given as they were determined afresh for children of London Elementary Schools.

Each child has to be tested individually, and under conditions most favourable to the removal of shyness and nervousness. The examiner must use his discretion respecting the point of the scale at which he should begin: the usual rule is to start with the group of tests just below the child's chronological age. If, however, there is a failure in any of those tests it is expedient to go back and try all the tests in the previous group. The examination should then be carried up the scale until the child fails in four or five consecutive tests.

Terman gives as the lowest permissible limit of thoroughness a range which starts at the year which yields only one failure, and ends with the year which yields only one success. To estimate the child's mental age the examiner should regard the age at which all tests are passed as the base age, and should add one-fifth of a year for every additional test belonging to any of the higher ages; or, where in the following revision there are more or less than five tests in the higher year, the corresponding reciprocal fraction.

It is now customary to state the final result in the form of an Intelligence Quotient, a method first suggested by the German psychologist, Stern. The Intelligence Quotient is found by dividing the mental age by the real age. If, for instance, a child's real age is eight and his mental age six, his intelligence quotient is ·75. If his mental age were ten his intelligence quotient would be 1·25.

Binet lays it down that the amount of retardation that determines a child as defective is two years when he is under nine, and three years when he is past his ninth birthday. Stated in terms of the intelligence quotient the border-line between normality and deficiency is somewhere about ·75. No cases, however, where the intelligence quotient falls between ·7 and ·8 are quite free from doubt.

BINET TESTS (BURT'S TRANSLATION AND REVISION)

Age Three

1. *Understanding Simple Commands.*

Instructions.—"Show me (point to, put your finger on)—

(1) your nose.
(2) your eyes.
(3) your mouth."

Evaluation.—All should be correctly performed; but free encouragement may first be given.

2. *Repeating Numbers.*

Instructions.—"I am going to say some numbers. Will you listen, and say them after me?"

	(For use only after failure in first set.)		
5	8	9	
3 7	6 4	7 2	(Age 3)
7 1 4	2 8 6	5 3 9	(Age 4)
3 6 8 1	5 7 4 9	8 5 2 6	(Age 5)
5 2 9 4 7	6 3 8 5 2	9 7 3 1 8	(Age 6)
2 5 0 6 3 4	5 7 3 9 1 6	4 9 5 8 2 7	(Age 9)
9 6 4 7 5 1 8	4 8 2 9 6 5 3	5 9 2 8 1 3 6	(Age 11)

Note: rate should be two per second; utterance should be without rhythm, emphasis or inflection. Do not tell the child if he is wrong. Do not repeat the same series. Merely give him

another chance with another series. Failure owing to interruption does not count. (While uttering the numbers or syllables, hold up the hand or finger to prevent the child starting to repeat before entire phrase or list has been completed. Drop the hand as a signal to child that you have finished and he is to repeat.)

Evaluation.—One correct repetition out of three trials counts as success. Note, therefore, the largest number the child can repeat. The age at which series of different lengths can be repeated is given in the last column above. The repetition of figures in their natural order, *e. g.* 9645678, should be noted as an instance of automatism.

3. *Naming Own Sex.*

Instructions.—"Are you a little boy or a little girl?" (for a boy). "Are you a little girl or a little boy?" (for a girl).

If child says "yes" or "no" or merely echoes part of the phrase, ask the two questions separately: "Are you a little girl?" "Are you a little boy?"

4. *Giving Surname.*

Instructions.—"What is your name?" If child merely gives Christian name, ask: "And what else? . . . Tommy what?"

Evaluation.—If child gives surname he has sometimes been known by *e. g.* stepfather, or mother (when illegitimate) record it as correct.

5. *Naming Simple Objects.*

Materials.—A penny, a closed knife, and a common kind of key.

Instructions.—"What is that?" or "What is this called?" showing each object successively.

Evaluation.—All three must be named, but slight errors, such as "money," "pennies," for "a penny," are allowable.

6. *Describing Pictures.*

Materials.—Binet's three pictures—chosen as containing people, and suggesting a story, and having a certain standardised difficulty.

There can, I think, be little doubt that pictures (1) better printed, (2) larger, (3) coloured, (4) representing actions in progress, (5) showing children, would be much more appropriate than Binet's original engravings. But these alone have been standardised.

Instructions.—"Look at this picture and tell me all about it." Binet's instructions are: "What is this?" and if the child says, "A picture," "Tell what you see there." It is better, however, to avoid leading phrases like "What can you see in it?" (which rather suggests enumeration) and "What are they doing?" (which suggests interpretation). Repeat instruction *once* for each picture, if there is no answer. Words of praise or encouragement alone may be added: "Isn't it a pretty picture? . . . Do you like it?" Or even

"That's right!" if the child is on the point of saying something but is withheld by shyness.

Evaluation of replies.—

Record the type of response given to the first picture; if doubtful, use the second and third and record the type of response most frequently given, *i. e.* employed for two pictures out of the three. Binet distinguishes three types of response corresponding to three stages of development.

A. Enumeration (E.). (Age 3.) *E. g.*—

i. "A man, boy."
ii. "There's an old man and a lady," etc. (mere list of objects or details).
iii. "I can see a room with a chair, a table, and a looking-glass, and there's a man and a sofa."

Two items at least should be enumerated. If the child only gives one, do not ask, "Anything else?" but proceed to another picture.

B. Description (D.). (Age 6.) (Phrases indicating actions or characteristics.)

i. "They're pulling a cart."
ii. "A man and a woman sitting on a seat." "An old man asleep."
iii. "A man standing on a bed and trying to look out of the window." "A man looking at himself in the glass."

C. Interpretation (I.). (Age 12.) (Goes beyond what is actually visible in the picture and mentions the situation or emotion it suggests.)

i. "They're moving," "they've a heavy load," "they can't pay their rent."

ii. "Miserable," "poor," "have no home," "the man is saying his prayers," "his daughter" or "wife" (looking after him, etc.).

iii. "A prisoner," "he wants to get out," "he's trying to see what's in the yard," "he is lonely" or "thinking," "a rag-picker," "a man in trouble," "a man on board ship."

Age Four

7. *Repeating Syllables.*

Instructions.—"Listen again, and say this after me." (The phrases should be pronounced deliberately and with expression. Begin with no. iii.; but if the child remains silent the examiner may give him first a shorter sentence (i. or ii.), and then apparently try iii. again.)—

i. (2 syllables). "Father."

ii. (4 syllables). "My hat and shoes."

iii. (6 syllables). "I am cold and hungry." (Age 4.)

iv. (8 syllables). "Here is the cloth; my hands are clean."

v. (10 syllables). "His name is Jack: he's such a naughty dog." (Age 5.)

vi. (12 syllables). "It is raining outside; but we can stay indoors."

vii. (14 syllables). "While Jack was doing his lessons, I caught a little mouse."

viii. (16 syllables). "We are going for a walk: Mary, let me see your pretty hat." (Age 7.)

. . . . . .

xiii. (26 syllables). "The other morning I saw in the street a little yellow dog. Little Maurice has spoilt his new apron." (Age 14.)

Evaluation.—Allow no error at all, except mispronunciations due to speech defects. (Binet's sentences appear to have been deliberately composed of two clauses. This seems unfortunate, as even an intelligent child may forget one. In translating them I have endeavoured to keep the general sense of the original, while making the phraseology more natural for a child.)

8. *Repeating Numbers.*

Instructions.—"Listen. Say these numbers after me"—

(For use only after failure in first set.)

7 1 4 2 8 6 5 3 9

Evaluation.—(See Test 2.)

9. *Counting Pennies.*

A. *Four Pennies.* (Age 4.)

Materials.—Four pennies placed in a row.

Instructions.—"Do you see these pennies? Count them, and tell me how many there are."

If the child at first answers at random, add: "Count them aloud, and point to each penny as you count it"; but do not demonstrate. (It may be of interest to see if the child can do it when shown: "Count like this: one, two——" touching the first two with the finger as each is counted. But do not use his answer for strict comparisons.)

Evaluation.—The first random answer does not count.

B. *Thirteen Pennies.* (Age 6.)

Instructions, etc., as before.

10. *Comparing Two Lines.*

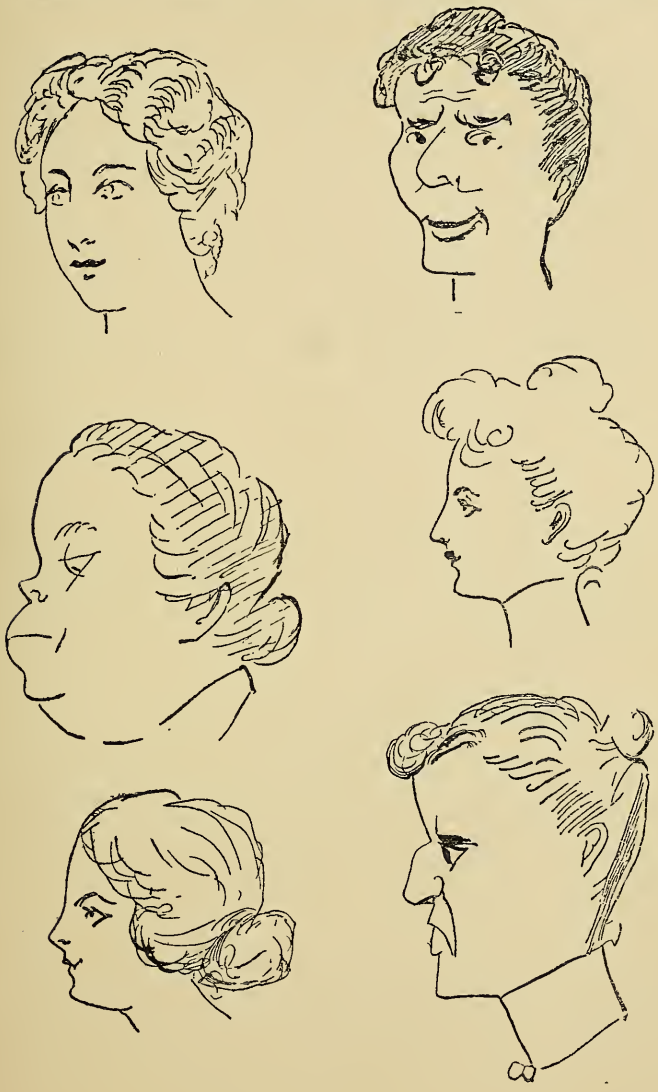
Materials.—Two parallel horizontal lines, 5 cm. and 6 cm. respectively, previously drawn in ink on a card or paper, the longer 3 cm. below the shorter, with its centre under that of the other.

Instructions.—"Do you see these lines? Tell me which is the longer?"

Evaluation.—No hesitation is allowed. (Some investigators allow the examiner to repeat the instruction: English children will often respond more readily to the injunction: "Put your finger on the long one." But Binet insists that the child shall not only perceive the difference, but also understand, without any further help, that the phrase "the longer" implies making a comparison.)

11. *Comparing Faces.*

Materials.—Binet's six faces. Show only two at a time.



Instructions.—"Which is the prettier of these two faces?" (If "prettier" seems not to be understood, "prettiest" or "Which do you like the best of these two ladies?" "Which is the nice one?" may evoke correct answers. But these should not be counted for purposes of strict comparisons.)

Method.—It is better to cover the lower pairs or pair while dealing with the first or second.

Evaluation.—All three comparisons must be correctly made. Repeat the questions once if necessary.

Age Five

12. *Performing Three Commissions.*

Materials.—Key, book, etc. Arrange the room while the child is carrying out one of the drawing or writing tests.

Instructions.—"Do you see this key? Go and put it on the table there. Then shut the door. And after that bring me the book on the chair near the door. Do you understand? First, put the key on the table, then shut the door, then bring me the book." (Note repetition of instructions. Do not let the child commence until this is completed.)

Evaluation.—All three commissions must be performed spontaneously *without* any further instruction or hint ("Well, and now?" "What have you forgotten?").

13. *Draws from Copy.*

A. *A Square.* (Age 5.)

Materials.—A square, each side measuring about 3 to 4 cm., drawn beforehand in ink, preferably on a card. Plain paper. Pen and ink (deliberately advised by Binet, making the task more difficult. Most American adapters recommend pencil.)

Instructions.—"I want you to copy this for me" (pointing to square). "Draw it here" (handing pen and paper). (If encouragement is needed: "What do you think this shape (picture) is? See if you can draw it." Do not use the word "square" yourself.)

Evaluation.—Passes if it can be recognised as an attempt at a square. If one side is twice the other, if the lines cross considerably at the corners or bend round without any angles, the drawing fails. The size does not matter. (Allow only one attempt.) Should take about one minute.

B. *A Rhombus or Diamond.* (Age 6.)

Materials.—A "diamond," about 7 cm. long and 4 cm. high, with sides 4 cm. long, drawn as before on a card. Paper, pen and ink.

Instructions and Evaluation.—As before. Binet apparently requires at least one pair of opposite angles to be fairly equal, at least one pair of adjacent sides to be fairly equal, and the vertical diameter to be long. Absolute parallelism of the opposite sides is not insisted upon. The pass-standard is thus considerably below what an uninstructed teacher would

be apt to accept as a satisfactory reproduction. (Reference should be made to his samples.)

14. *Repeating Syllables.*

(10 syllables.) "His name is Jack: he's such a naughty dog."

Instructions and Evaluation.—(See Test 7, no. v.)

15. *Giving Age.*

Instruction.—"How old are you?"

Evaluation.—Child should give his age in years, last birthday. Note: children very often say "seven" when they mean "getting on for seven." Hence, if the first answer is wrong, ask specifically "How old were you last birthday?" Parents also often give an infant and a child about to leave school an age above the true one: and dull children (except when about to leave) an age below the real one. The child's answer should be accepted if it corresponds with what it has commonly or recently been told. Do not, therefore, insist too rigidly on the age given by the birth certificate or the register.

16. *Distinguishing Morning and Afternoon.*

Instructions.—"Is it morning or afternoon now?" (in the morning). Or, "Is it afternoon or morning now?" (in the afternoon).

Evaluation.—Repeat the question if there is any possibility of the child having merely echoed one of the words thoughtlessly. (Asking "Have you had your dinner yet?" elicits answers interesting to compare with the above.)

17. *Naming Four Primary Colours.*

Materials.—Four oblong pieces of paper, 2 × 6 cm., coloured bright red, yellow, blue and green, and gummed beneath one another on a card.

Instructions.—"What colour is this?" pointing to each in turn.

Evaluation.—No error is allowed. Should take about 6 seconds. But the time limit does not appear to be strictly enforced.

18. *Repeating Numbers.*

Instructions.—"Listen. Say these numbers after me."

(For use only after failure in first set.)

3 6 8 1 5 7 4 9 8 5 2 6

Evaluation.—(See Test 2.)

19. *Comparing Two Weights.*

Materials.—Four small similar boxes (about 1·5 × 2·5 × 3·5 cm.) weighing 3, 12, 6 and 15 grammes.

Instructions.—"You see these boxes (showing first the pair weighing 3 and 12 grammes placed 5 or 6 cm. apart). Tell me which is the heavier."

If the child merely points, add, without any gesture: "Take them in your hands and weigh them." (English children respond better to the instruction "Lift them" or "Feel them." "And give me the heavy one." But do not use the modification if strict comparability is required. Kuhlman and the Stanford Revision allow a demonstration: Binet and Yerkes prohibit it.) In any

case do not put them in his hands. If he merely lifts one, or both together, do not correct him. If there is any doubt with the first pair, repeat the experiment with the second pair (6 and 15 grammes); and then with the first pair again. (If the child fails to understand, it is interesting to put them successively into his hand, and ask, "Which is the heavier?" But his response in this case does not count.) For the second and third trials use the 6 and 15 gramme weights and then the first pair again.

Evaluation.—All three trials must be correct: if any doubt continue repetitions.

20. *Giving Number of Fingers.*

Instructions.—"How many fingers have you on your right hand?" "And how many on your left hand?" . . . "How many does that make on both hands altogether?"

Evaluation.—The replies must be made without stopping to count: and all three questions must be correctly answered.

Age Six

21. *Counting Pennies.*

A. *Thirteen Pennies.*

Materials.—Thirteen pennies placed in a row.

Instructions and Evaluation.—(See Test 9, B.)

22. *Drawing from Copy.*

B. *A Rhombus or Diamond.*

Instructions and Evaluation.—(See Test 13, B.)

23. *Transcription.*

Materials.—"See little Paul" written in a bold, copy-book handwriting on a card or sheet of paper. Paper, pen and ink.

Instructions.—"Will you copy that for me?"

Evaluation.—The test is passed if the copy is sufficiently legible to be read by a person who did not know what was to be written.

24. *Naming Days of the Week.*

Instructions.—"Can you tell me what are the days of the week?"

Evaluation.—The days must be named in order without error or hesitation, in 10 seconds.

25. *Naming Coins.*

Materials.—Nine coins, all placed in a row on the table with the head upwards: similar coins should not be adjacent. Order upon table: 1*s.*, 1*d.*, 6*d.*, ½*d.*, 2*s.*, ¼*d.*, £1, 2*s.* 6*d.*, 10*s.* Order of difficulty: 1*d.*, ½*d.*, ¼*d.*, 1*s.*, 6*d.*, 2*s.*, £1, 10*s.*, 2*s.* 6*d.* (3*d.*, 5*s.*, 4*s.*). (One pound and ten-shilling notes must be allowed.)

Instructions.—Ask "What is this?" pointing to each in succession. Neither examiner nor child should handle them or turn them over.

Evaluation.

A. *Four Commonest Coins.* (Age 6.)

(1*s.*, 6*d.*, 1*d.*, ½*d.*) No error allowed.

B. *Nine Commonest Coins.* (Age 9.)

All should be named correctly in 40 seconds. If an error is attributable to passing confusion, Binet

allows a second trial of the whole series after a few minutes. (An interesting variant is to ask what coins there are larger than a shilling, before passing to B.)

26. *Reconstructing Divided Oblong.*

Materials.—Two cards (4·5 × 7·5 cm.), one intact, the other divided along one diagonal into two equal triangles. Place the triangles so that the longest sides are at right angles, but do not face towards each other.

Instructions.—" One of my cards has been cut in two; can you put the pieces together again, to make a whole one, like this? "

If the child merely looks at the cards without touching them, say: " Move them about and see if you can fit them together," and, if necessary, place one in his hand.

See that the child does not turn one triangle over. (Before cutting the card, black one side all over. This does not appear to alter the difficulty of the test, but prevents turning over.) If the child makes a wrong combination and appeals for judgment, give no opinion. Remain silent, or say merely, " What do *you* think? "

27. *Defining Concrete Terms.*

Instructions.—" What is—

(1) a fork?
(2) a table?
(3) a chair?
(4) a horse?
(5) a mother? "

(A word commonly used by other investigators is "kitten," or "cat": "fork" is unfortunately a difficult word to begin with.) For shy or silent children: "You know what a fork is, don't you? Well, tell me what it is—what is a fork?" "You have seen a horse, haven't you? Tell me what a horse is." The instructions may be repeated, but use no other form of words. (Give the child a minute to reply in.)

Evaluation of Replies.—The character of three replies out of five determines the value of the test. The variations in the age assignments of definition superior to use depend largely on the inclusion of such replies as ii., (1), (3), (4) and (7) under i. rather than under ii.

Note "U" or "G" according as child defines—

i. In terms of *Use*. (Age 6.)

(1) *What* you eat *with*.
(2) *Something* to have your dinner on: where the plates are put.
(3) It draws a cart.
(4) She takes care of the babies.

ii. In terms *superior to Use* (by Genus with or without Differentia, or by Description). (Age 10.)

(1) A *thing* to sit on: *something* that you eat with. ("Thing" and "something," however, are not accepted for "horse" or "mother.")
(2) An *instrument*.
(3) It has four legs: it's silver.

(4) A piece of wood, part of the furniture.
(5) An animal.
(6) A lady.
(7) One who cooks our dinners.

(Note: if a child uses "thing" or "something" for "chair," I then give "mother" and "horse," otherwise even a bright child, having given "thing" for "chair," "table" and "fork" without correction is apt from sheer inertia to offer "thing" as the genus of "horse.")

iii. Merely repeating the word, or pointing to the object is marked a failure, without (apparently) giving the child a further chance.

28. *Repeating Numbers.*

Instructions.—"Listen. Say these numbers after me."

(For use only after failure in first set.)

5 2 9 4 7 6 3 8 5 2 9 7 3 1 8

Evaluation.—(See Test 2.)

29. *Describing Pictures.*

Description (D.) (Phrases indicating actions and characteristics.)

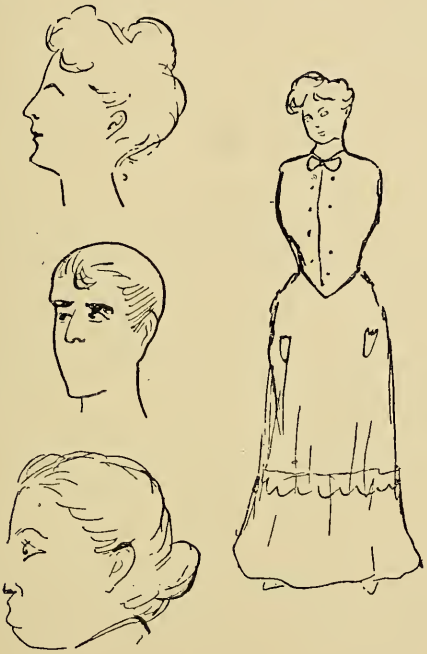
Instructions and Evaluation.—(See Test 6, B.)

30. *Distinguishing Right and Left.*

Instructions.—"Which is your right hand?" . . . "Which is your left ear?"

Evaluation.—The child must perform both correctly without any kind of help. Hesitation and

self-correction (without any hint) are allowed: if by a slip the child shows his left hand or right ear, the experimenter waits a moment for a spontaneous correction, which is allowed to pass, but his manner of waiting should not suggest that the first action was wrong.



Age Seven

31. *Recognising Missing Features.*

Materials.—Binet's four pictures of faces without mouth, nose, eye, and of body without arms.

Instructions.—"Look at this lady's face. Can

you tell me what has been left out?" (Begin with face without mouth.)

If the child says, "Her body," reply, "Oh, I was only trying to draw her face. What must I put in to finish the drawing of her face?" ("What have I forgotten in drawing her face?")

Evaluation.—Three correct answers out of four are required.

32. *Adding Three Pennies and Three Halfpennies.* (Concrete.)

Materials.—Three pennies and three halfpennies set out in a row. (American investigators commonly use stamps, but to many children the value of these is unfamiliar.)

Instructions.—"Will you count this money for me, and tell me how much there is altogether?"

Evaluation.—No error and no repetition of the instruction is allowed. The test should be done in 8 to 10 seconds. "It is useless to wait 15 seconds."

33. *Stating Differences between Concrete Objects.*

Instructions.—"You know what wood is, don't you? . . . And you know what glass is? . . . They are not the same, are they? . . . In what way are they not the same?" (I would add, if child hesitates: "They are different, are they not? Well, do you think you can tell me what the difference is? . . . How can you tell 'glass' from 'wood'?")

The following words are suggested by Binet—

i. fly, butterfly;
ii. wood, glass;
iii. paper, cardboard.

Evaluation.—Two out of three statements must be correct. Any true difference will pass, though trivial. But if the child repeats the same difference, *e. g.* "It is larger," it is insufficient. (Ask, "In what other way are they not the same?") Often a child takes a minute: but if he takes longer than 2 minutes for all he fails.

34. *Repeating Syllables.*

(16 syllables.) "We are going for a walk: Mary, let me see your pretty hat."

Instructions and Evaluation.—(See Test 7, no. viii.)

35. *Writing from Dictation.*

Materials.—Pen, ink, paper.

Instructions.—"Will you write this down for me on this piece of paper?"

"The pretty little girls."

Evaluation.—The writing (? and spelling) must be sufficiently legible and accurate to be read by a person who did not know what was to be written.

Age Eight

36. *Reading and Reproduction.*

Material.—Translation of Binet's passage, printed or typed, with English place-names and money-values substituted for the French.

Three / Houses / on Fire./
London, / September 5th. / A huge fire / last night/ burnt down / three houses in the middle of the city. / Seventeen families / now have no homes. / The loss is more than 15,000 pounds. / A young barber,/ who saved /a baby / in its cradle, / was badly / hurt/ about the hands./

Instructions.—"Will you read this for me, please?" Two seconds after reading is finished, remove the passage, and say: "Tell me what you have been reading about."

Evaluation.—Each correct phrase or word as indicated above constituted one item.

A. *Recalls two items.* (Age 8.)
B. *Recalls six items.* (Age 9.)

37. *Answering Easy Questions.*

Instructions.—"Tell me this—

(1) Suppose you are going somewhere by train. What must you do if you miss your train?
(2) Suppose one of the other boys (girls) hit you by accident—without meaning to. What should you do?
(3) Tell me what you ought to do if you broke something that belonged to somebody else?"

If no answer is given, repeat the question as usual, not sternly, but pleasantly, prefixing: "Did you catch what I said?" Do not vary the wording.

Evaluation.—Two out of three must be answered satisfactorily.

(1) *Satisfactory answers.*—"Wait for another"—"Take the next."
Unsatisfactory answers.—"Go home again"—"Run after it"—"Try not to miss it."

(2) *Satisfactory answers.*—"Do nothing"—"Forgive him."
Unsatisfactory answers.—"Tell teacher"—"Hit him back."

(3) *Satisfactory Answers.*—"Pay for it"—"Own up"—"Buy another"—"Ask to be forgiven"—"Say I was sorry."
Unsatisfactory answers.—"I should cry"—"Hide it."

38. *Counting Backwards* 20–0.

Instructions.—"You can count, can't you—1, 2, 3, and so on? Now do you think you could count backwards? Start at 20, and go on till you reach 1." (If the child does not understand:) "Count like this: 20, 19, 18" (proceed no further).

(Yerkes suggests that the experimenter always count from 25 to 21, and then pause for the examinee to continue.)

Evaluation.—One error (either of omission or inversion) only is permitted (Binet allows 20 seconds). The child who thinks out the numbers by counting up from 1 each time fails.

39. *Giving Full Date.*

Instructions.—"What is the date to-day?" (If the word "date" is not understood, ask in detail: "What day of the week is it to-day?" "What month is it?" "Do you know what day of the month?" [1st, 2nd, 3rd, or] "what number?" "And what is the year—nineteen what?")

Evaluation.—All four items must be correctly given: but an error of three days either way is allowable for the day of the month (unless that involves an error in naming the month).

40. *Giving Change.*

Materials.—The current coins $\frac{1}{4}d.$, $\frac{1}{2}d.$, $1d.$, $6d.$, $2s.$, $2s.\ 6d.$, $10s.$, £1 and $1s.$, and in addition three pennies and the three halfpennies. The five boxes used for the weights.

The shilling is kept by the experimenter to pay for the box.

The rest, with the boxes, are placed near the child.

Instructions.—"Now, shall we play shop for a change? You shall be the shopkeeper. Here are some boxes for you to sell; and here is your money. See how rich you are! Now, will you sell me one of your boxes, please? How much are they each? Twopence, shall we say? Well, here is a shilling. Can you give me the right change, please?" (The examiner holds out his hand for the money.)

Evaluation.—The child must actually hand over the right amount (sixpence and fourpence in pennies or halfpennies): merely stating it correctly does not count.

Age Nine

41. *Repeating Numbers.*

Instructions.—" Listen. Say these numbers after me "—

(For use only after failure in first set.)

2 5 0 6 3 4 5 7 3 9 1 6 4 9 5 8 2 7

Evaluation.—(See Test 2.)

42. *Naming Months.*

Instructions.—" Can you tell me what are the months of the year ? "

Evaluation.—Binet allows 15 seconds, and one error.

43. *Naming Coins.*

Nine Commonest Coins.

Instructions and Evaluation.—(See Test 25, B.)

44. *Reading and Reproduction.*

Recalls Six Items.

Instructions and Evaluation.—(See Test 36, B.)

45. *Defining Concrete Terms.*

Superior to Use (G.).

Instructions and Evaluation.—(See Test 27, no. ii.).

46. *Arranging Five Weights in Order.*

Materials.—Five boxes, identical in colour, shape and size (about 1·5 × 2·5 × 3·5 cm.), and loaded with shot and cotton wool to weigh 3, 6, 9 and

15 grammes, without rattling. The key letters, B, I, N, E, T, may be written in order on the bottom of the boxes.

Instructions.—"Do you see these boxes? They all look the same, don't they? But they don't weigh the same. Some are heavy, and some are light. I want you to find the heaviest of all, and put it here. Then find the one which is nearly as heavy, and place it next. Then the one which is still less heavy; then the one which is lighter still; and, last of all, the one which is lightest of all, here."

Allow three trials if necessary, mixing the boxes up first.

Evaluation.—The arrangement must be absolutely correct in two out of three trials, and the whole accomplished in 3 minutes. It is of special interest to record the subject's actual arrangement.

47. *Sentence Building with Three Words.*

Materials.—Paper, pen and ink: and a card with "London, river, money" written on it.

Instructions.—"I want you to make up a sentence for me with these three words in: London, river, money."

Instead of "London" it is often customary to employ the nearest town that is on a river. (Most American investigators, following Goddard, conduct this test orally.)

Evaluation.

A. *Two distinct ideas or sentences* given (indicates mental age of 10). "London has money and rivers."

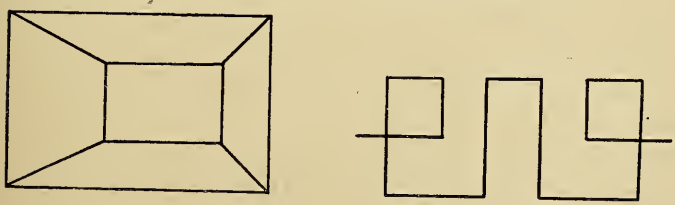
B. *One idea or sentence* given (indicates mental age of 11), *e. g.* "In the river at London I found some money." A set of sentences in which the thought is well co-ordinated into a unitary story or description passes.

C. Three distinct ideas or sentences constituted a failure. *E. g.* "London is a town. There is a big river. Some people have money."

Enter "1," "2" or "3" according to number of sentences given, and note time. At least three-quarters of the test should be written within a minute. Binet states that this is one of the rare cases in which a child may succeed by having heard of the test from another child. If there is any likelihood of this, ask at the outset: "What do you think I have been asking the others to do with these words?" and substitute others, if necessary. The child may guess the test from school exercises.

48. *Drawing Two Designs from Memory.*

Materials.—Binet's two designs, drawn previously on a single card or sheet, kept out of sight until required. A pencil and plain paper.



Binet's Test No. 48.

Instructions.—" I am going to show you two easy drawings. I want you to look at them very carefully until I take them away. Then, after I have turned them over, see if you can draw them both from memory on this paper. You will only see them for a very few seconds. Now look at them both carefully first of all. Ready? Now!" The drawings are held steadily in front of the child, the truncated prism on the left, for 10 seconds (see that the child does not imagine he has to copy them at once), and then taken away and turned face downwards. " Now try and draw them for me here."

Evaluation.—The whole of one and a half of the other must be reproduced fairly exactly. No second attempt is allowed. Neatness of drawing does not count. The examiner must be careful not to interpret " fair exactness " more strictly for older than for younger children. The standard accepted in the case of this test is thus far below what the uninstructed teacher would accept as a satisfactory reproduction.

Age Eleven

49. *Explaining Absurdities.*

Instructions.—" Listen carefully to what I am going to say. There is something in it that is really quite silly and impossible. See if you can tell me what is wrong."

i. " ' One day, a man fell off his bicycle on to his head and was killed instantly. He was taken to the hospital and they say he will never get better.' What is there silly in that? "

ii. " 'Once the body of a poor girl was found in a wood, cut into eighteen pieces. They say that she killed herself.' What is silly in that?"

iii. " 'Yesterday there was a railway accident. But the newspaper says it is not a serious one, as only forty-eight people were killed.' What is silly in that?"

iv. " 'I have three brothers—Jack, Tom and myself.' What is silly in that?" (Female examiners must preface this with, "A boy said to me, etc.," or else substitute, "I have three sisters, Jane, Mary and myself.")

v. " 'A man once said, "If I should ever grow desperate and kill myself, I shall not choose a Friday to do it on, for Friday is an unlucky day, and would bring me bad luck." ' What is foolish in what the man said?"

Evaluation.—Three absurdities should be detected out of five. If a child's first statement is not clear (*e.g.* "myself is silly" in answer to iv.), say, "Explain what you mean." Otherwise, give him no second chance.

(i.) *Correct:* "He couldn't get well if he was already dead." "First you said he was dead, and then you said he wouldn't get well again."

Incorrect: "They ought to have taken him to the mortuary." "If he fell off his bicycle, he wouldn't fall on his head."

(ii.) *Correct:* "You can't cut yourself into eighteen pieces." "If she killed herself she couldn't cut herself up."

(iii.) *Correct:* "It must have been serious if forty-eight were killed, or if anybody was killed." "If it wasn't serious only one or two would have been killed." "Forty-eight isn't serious in war-time."

Incorrect: "Forty-eight people couldn't be killed in a railway accident."

(iv.) *Correct:* "You have only two." "You are not your own brother." "You shouldn't count yourself."

Incorrect: "You should put yourself last."

(v.) *Correct:* "If he killed himself, the day wouldn't matter." "He couldn't have bad luck if he was dead." "If he was desperate, he wouldn't wait till Friday."

Incorrect: "He is silly to believe in bad luck." "Friday isn't different ~~to~~ any other day."

50. *Answering Difficult Questions.*

Instructions.—"Can you tell me this?"

(1) "What should you do if you found you were late on your way to school?"

(2) "Suppose a boy does something that is unkind: why do we forgive him more readily if he was angry than if he was not angry?"

(3) "If some one asked what you thought of a boy (or girl) whom you did not know very well, what should you say?"

(4) "Why should we judge a person by what he does, and not by what he says?"

(5) "Suppose you were going to undertake something very important: what should you do first of all?"

Repeat a question once, if necessary, but do not vary the wording.

Evaluation.—Allow 20 seconds for reflection on each question. Three out of five must be answered satisfactorily.

(1) *Satisfactory:* "Hurry" or "Run" ("Go straight to school" may be accepted if it appears that the child sometimes plays or carries out errands on its way).
Unsatisfactory: "Get the stick." "Leave earlier." "Get up sooner next time." "Ring the bell." "Get a note to excuse me." By convention, anything not embodying the idea of hurrying.

(2) *Satisfactory:* "Because he didn't know what he was doing." "Because he'd be sorry afterwards." Anything suggesting that anger may constitute an excuse, however badly expressed.
Unsatisfactory: "He oughtn't to get angry." Anything suggesting disapproval of anger.

(3) *Satisfactory:* "I could not say anything." "I could not tell him without finding out." "I should say, 'I do not know.'"
Unsatisfactory: "I should have to ask." "Say I did not know his name." Usually unintelligible.

(4) *Satisfactory:* "You can rely on his actions, but not on what he says." "Because he might not always speak the truth." "Actions speak louder than words."
Unsatisfactory: Usually unintelligible. "Because you can't tell." "You ought to speak the truth."

(5) *Satisfactory:* "Think it over." "Ask some one about it." "Prepare for it."
Unsatisfactory: Usually unintelligible. "Not do it."

(Some say "Tidy myself," "Put on a clean collar;" in that case make it clear that you mean *doing* something important, not *going* somewhere important.)

51. *Repeating Numbers.*

Instructions.—"Listen. Say these numbers after me."

(For use only after failure in first set.)

9647518 4829653 5928136

Evaluation.—(See Test 2.)

52. *Giving Sixty Words in Three Minutes.*

Instructions.—"I want you to give me as many words as you possibly can in 3 minutes. Keep saying words like this till I stop you: school, teacher, board, boy, girl, and so on. Some children can give more than 200. Are you ready? Now start." When he stops encourage him immediately by saying: "Very good. Keep on."

Evaluation.—Sixty words must be given, exclusive of repetitions. If the child gives sentences, start him again, saying: "You must give separate words." Be careful (1) to note the time, if possible with a second hand (2) to count the words, entering a stroke or other mark for each, and calculating the total. It is interesting to record the key-words of the child's various topics; it is seldom possible to put down all words.

53. *Sentence Building with Three Words.*

Materials.—Paper, pen and ink, and a card with "London, river, money," written on it.

One Idea or Sentence.

Evaluation.—(See Test 47, B.)

Age Twelve

54. *Giving Three Words to Rhyme.*

Instructions.—"Do you know what a rhyme is? When two words end the same way, we call them rhymes. 'Jill' rhymes with 'hill' because they both end in 'ill.' Do you understand? Now can you give me three words which rhyme with 'obey'?"

Evaluation.—The child must give three genuine words that rhyme in 1 minute. Binet's instructions to the child ask for "other words" or "all the words." It saves time to specify three to the child. If the child gives nothing, or has not given enough, urge him by saying "what (else) rhymes with 'obey.'" Apparently "disobey" may be accepted as one of the three.

55. *Rearranging Mixed Sentences.*

Materials.—Three cards containing the following words—

i. a defends dog good his bravely master;
ii. my have teacher I the correct asked paper to;
iii. home we early our in country left visit the to friends.

Instructions.—"Put these words in order, and find out the sentence which they make."

Evaluation.—Two correct solutions must be given out of three. Only 1 minute is allowed for each.

Correct solutions are—

i. "A good dog defends his master bravely."
"A dog defends his good master bravely."
("A master defends his good dog bravely," is, according to Binet, "poor" and apparently incorrect).

ii. "I have asked my (the) teacher to correct the (my) paper."
("I asked my teacher to have the paper correct" is, presumably, incorrect.)

iii. "We left home early to visit our friends in the country."
("We left our friends in the country to visit our home early," and other sensible variants are presumably correct.)

56. *Describing Pictures.*

Interpretation.—(Goes beyond what is actually visible in the picture, and mentions the situation or emotion it suggests.)

(See Test 6, C.)

Age Thirteen

57. *Resisting Suggestion.*

Materials.—A book of six leaves with two lines drawn in the same straight line on one page in opening. The lengths must be as follows—

	1st line.	2nd line.
1st page	4 cm.	5 cm.
2nd page	5 cm.	6 cm.
3rd page	6 cm.	7 cm.
4th page	7 cm.	7 cm.
5th page	7 cm.	7 cm.
6th page	7 cm.	7 cm.

Instructions.—For the first three pages : " Which is the longer of those two lines ? " for the last three, without changing the tone : " And these ? "

Evaluation. — Note whether child's judgments are right or wrong in each case, especially with reference to the last three. The child succeeds if he judges two out of three equal lines to be equal.

58. *Solving Circumstantial Problems.*

Instructions.—" Can you guess the answer to this riddle ?

i. " One day a woman, walking in Epping Forest, stopped still, terribly frightened. Then she hurried to the nearest police station, and told the policeman she had just seen, hanging from the branch of a tree, a —— well, what do you think it was she saw ? "

ii. " My next-door neighbour has had three visitors. First, a doctor called ; then a lawyer ; and then a clergyman. What do you think has been happening there ? "

Evaluation.—Both questions must be correctly answered.

i. *Correct:* Replies must contain the idea of some one hanged. If the child answers "a man," "a dead person," therefore, ask "How did he get up in the tree?"
Incorrect: "A bird." "Some one robbing a nest."
ii. *Correct:* "Some one is dying," "is very ill." (Severe illness can be inferred from the visit of the doctor alone. But Binet apparently accepts it, without inquiring whether the child knows the object of the other visitors.)

Age Fourteen

59. *Repeating Syllables.*

(26 syllables.) "The other morning I saw in the street a little yellow dog." "Little Maurice has spoilt his new apron."

Evaluation.—(See Test 7, no. xiii.)

60. *Defining Abstract Terms.*

Instructions.—"Can you tell me this? What is meant—

i. by *kindness?*
ii. by *justice?*
iii. by *charity?*"

(For "charity" some investigators have substituted "obedience": but this changes the difficulty of the test.)

Evaluation.—Two must be correctly defined out of three. Correct definitions—

For (i) contain the idea of an instance of affection, tenderness, politeness or consideration to others. ("Being polite or good to others" is correct. "Being kind," "doing something good," are inadequate.)

For (ii) contain the idea of treating people according to their merits, or of protecting the innocent and their interests, or of punishing the guilty. *E. g.* "when you punish wicked people," "playing fair."

For (iii) contain the ideas of (*a*) poor or unfortunate people, and (*b*) of showing kindness. *E. g.* "when you give poor people some money," "giving alms."

Age Fifteen

61. *Drawing from Imagination the Cuts in a Folded Paper.*

Materials.—Two sheets of paper about 6 inches square. A pencil. One sheet is folded in four like a letter ready for an envelope. In the middle of the edge which presents but a single fold, a small triangular notch, about 1 cm. deep, is drawn.

Instructions.—"Here is a sheet of paper that has been folded across and then folded again. Suppose now I cut out a notch just here. When the paper is unfolded again, what would it look like? Will you show me on this piece how and where it would be cut?"

Evaluation.—Two diamond-shaped holes should be drawn in a line with each other, each in the middle of one half of the paper.

62. *Giving Differences between Abstract Terms.*

Instructions.—" What is the difference between—

i. pleasure and happiness ?
ii. poverty and misery ?
iii. evolution and revolution ? "

The words suggested by Binet (*Bulletin*) are—

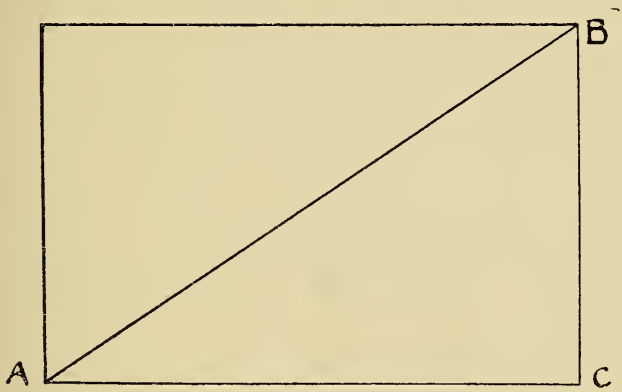
i. paresse, oisiveté ;
ii. événement, avènement ;
iii. évolution, révolution : and in 1908 scale also—
iv. plaisir, bonheur ;
v. orgeuil, prétention.

Evaluation.—Two out of three must be correctly answered. Good replies should bring out an opposition or antithesis between the differentiating ideas. *E. g.* (i) " happiness " is superior to or more general than " pleasure " ; (ii) should contrast having little money with being in misery or pain ; (iii) should contrast slow change with sudden change. But Binet admits mere differences ; *e. g.* evolution is the movement of troops, revolution is an insurrection.

63. *Drawing the Displaced Triangle.*

Materials.—Paper and pencil for drawing. A card about 10 × 15 cm., cut across the diagonal, as used for divided card test. The card is first laid

on the table before the subject with the cut edges touching.



Instructions.—" Look carefully at the lower piece of this card. Suppose I turn it over and lay this edge (A–C) along this edge (A–B of the upper triangle), and suppose that this corner (C) is placed just at this point (B), what would it all look like? Now I am going to take the piece away (remove lower triangle from view). Imagine it placed as I told you, and draw its shape in the proper position. Begin by drawing the outline of the top triangle."

Evaluation.—The essential points are: (i) A C B must be preserved as a right angle; (ii) A C must be made shorter than A B.

64. *Summarising Hervieu's Reflections on Life.*

Instructions.—" Attend carefully to what I am going to read you. When I have finished I shall want you to tell me the *meaning* of what I read in

your own words. The exact words that I use do not matter. Listen—

'Many opinions have been given on the value of life. Some say it is good, others say it is bad. It would be truer to say that it is just medium. For, on the one hand, the happiness it brings us is never so great as we ourselves should like, and, on the other hand, the misfortunes it brings us are never so great as our enemies would want us to have. It is the intermediate nature of life that makes it fair, or, at least, prevents it from being altogether unfair.'

Now see if you can give me, in your own words, the sense of what I have just read to you."

Evaluation.—The central thought must be understood and these three ideas reproduced: (1) Life is neither good nor bad, but medium, for (2) it is not so good as we wish, but (3) better than what others wish for us. The terms and expressions matter little.

65. *Giving the Differences between President and King.*

Instructions.—"There are three chief differences between a King and a President of a Republic. Can you tell me what they are?" "Can you think of any of them?"

Evaluation.—Two of the following differences, apparently, must be given. (Some consider Binet required all of the first three.)

i. A King inherits his crown: a President is elected.

ii. A King is king for life : a President's term of office is limited.
iii. The powers of a King are greater than those of a President.
iv. A King is not directly responsible to the people : a President is. (Added by Melville in place of iii.)

(This test is obviously more suited to French and American children than to English. The third difference is hardly true of an English king.)

CHAPTER V

BURT'S REASONING TESTS

A TEACHER who has studied Binet's tests will at once perceive that they are unsuitable for the upper standards. They were not devised for the discovery of the bright children, but for the detection of the dull. If the aim is to select supernormal children to whom scholarships might be awarded, a more difficult series of tests—a series that appeals more exclusively to the higher mental functions—is necessary. The tests recently published by Mr. Cyril Burt and printed below were designed to meet that need. A full description of the circumstances under which the tests were drawn up and standardised, together with an account of the development of reasoning in school-children, will be found in an article by Mr. Burt in the *Journal of Experimental Pedagogy* for June 1919 and December 1919.

To quote from that article: "The test-questions are intended to be given to each child individually and orally. In my own experiments each problem was type-written upon a separate card; a fresh statement commenced on a fresh line; and by means of indentation and spacing, question and premises were distinguished from each other. A card is handed to the child with the following instructions; 'Will you read this little puzzle?

There is an easy question at the end. When you have read the question, read carefully again what is printed above, and try whether you can think of the answer.' The younger and duller children should read the test-questions aloud; and with the youngest and dullest of all, the examiner should read the questions with or to the child. Children of higher levels (Standard III) need only read aloud the first few questions. Any child who is unable to read a particular word or to comprehend its meaning should be freely helped. The graver incongruities between difficulty of phrase and difficulty of logic have been eliminated. In a perfectly revised list they should never occur. A bright young child is occasionally puzzled by such words as 'sub-tropical,' and 'emotion,' although competent to follow the reasoning. When it is clear that the child understands his task, he should be left quietly with the card, forgetful, if possible, of the examiner's presence. The emotional confusion, the 'examination paralysis,' that so commonly embarrasses an oral interview is by this means largely avoided. When the child gives an answer, it is invariably received with a word of praise, whether right or wrong; and the child is asked to give his reason.

"One mark is given for each test correctly answered and correctly reasoned. When necessary, the child may be given additional trials, not exceeding three in all for any one test. But for each unsuccessful attempt a quarter of a mark is deducted. A fraction—as a rule, a quarter, a half, or three-quarters respectively—is also deducted for an ill-expressed reason, an inadequate reason, or no reason at all.

In the cross-examination as to reasons lies the most valuable part of the test. The examiner gleans considerable information, not only about the knowledge and intellectual procedure of the child, but also about its temperament and disposition so far as they affect his intellectual efficiency. In the final estimate of the child he would take into account both these broader observations, and in particular the speed with which the child has worked. In the actual marking no allowance is made for such latter factors; nor is any time-limit assigned. Could scores be corrected on the basis of the general impressions incidentally gained, the correlations with ability, high as they are, would be still further raised.

"For a child to work steadily through a series of fifty reasoning tests until he breaks down would, at any rate for the brighter and older children, be a slow and fatiguing process. A short series has therefore been constructed by selecting every third test in the full series. The short list thus contains only seventeen questions, two for each age except the first, which has three.

"In the full list appended these selected questions are marked with asterisks. They have been more carefully chosen, more extensively used, and more thoroughly revised. For practical purposes, indeed, the short list will be sufficient, since this allows a rough and rapid determination of mental age. Where, however, it is required to obtain a more exact estimate of a child whose mental level is already approximately known—for example, in testing children within the same school standard—the full list is indispensable, since with the short

list no member of a fairly homogeneous class could be expected to differ from the others by more than one or two marks.

" Children should be tested with the short list first. Even the oldest and brightest should begin with the easiest test. They should be carried through the series until they have broken down with three consecutive tests. The supplementary questions should be given subsequently, and upon a different day. Here it will be expedient to start, not at the beginning of the series, but about four tests below the level of the first serious failure made on the short list; and the child should be carried through until he breaks down on at least six tests consecutively."

BURT'S GRADED REASONING TESTS.

Seven Years

*1. Tom runs faster than Jim: Jack runs slower than Jim. Who is the slowest—Jim, Jack or Tom?

2. All wall-flowers have four petals: this flower has three petals. Is this a wall-flower?

3. It looks like rain: but I shall stay indoors. Shall I want an umbrella to-day?

*4. Kate is cleverer than May: May is cleverer than Jane. Who is the cleverest—Jane, Kate or May?

5. It is Sunday; and on a Sunday afternoon Ada usually takes the baby out, or goes by herself to the pictures, or walks over to see her aunt, or else goes by tram to the cemetery. To-day she has no money with her: and the baby is asleep upstairs. Where do you think she has probably gone?

6. Tom said to his sisters: " Some of my flowers

are buttercups." His sisters knew that all buttercups are yellow. So Mary said: "All your flowers must be yellow." Grace said: "Some of your flowers must be yellow." And Rose said: "None of your flowers are yellow." Which girl was right?

*7. I have bought the following Christmas presents: a pipe, a blouse, some music, a box of cigarettes, a bracelet, a toy engine, a bat, a book, a doll, a walking-stick and an umbrella. My brother is eighteen: he does not smoke, nor play cricket, nor play the piano. I want to give the walking-stick to my father and the umbrella to my mother. Which of the above shall I give to my brother?

Eight Years

8. All great men work hard and long every day: Sir John Smith worked three hours a day. Was Sir John Smith a great man?

9. Peter has a half-holiday on Wednesdays and Saturdays, and a whole holiday on Sunday. I am at work all day, except on Monday, Wednesday, Friday and Sunday. I want to take Peter to the tailor's to buy a new suit. Which afternoon could we go together?

*10. I don't like sea voyages; and I don't like the seaside. I must spend Easter either in France, or among the Scottish Hills, or on the South Coast. Which shall it be?

11. Ethel has twice as many apples as John: Lucy has half as many as John: Lucy has ten. How many has Ethel?

12. Edith is fairer than Olive, but she is darker than Lily. Who is darker—Olive or Lily?

*13. The person who stole Brown's purse was neither dark, nor tall, nor clean-shaven. The only persons in the room at the time were: (1) Jones, who is short, dark and clean-shaven. (2) Smith, who is fair, short and bearded. (3) Grant, who is dark, tall, but not clean-shaven. Who stole Brown's purse?

Nine Years

14. C is smaller than B: B is smaller than A. Is A greater than C?

15. A burglar entered my room at the Hotel Splendid last night. The windows were all securely fastened on the inside, and the fastenings and the window-panes are undisturbed. The opening up the chimney is only nine inches square. The door opening into the main corridor was locked and the key left on the outside. The ceilings, walls and floor have no other openings, either secret or forced, through which he could have entered. How did he get in?

*16. Three boys are sitting in a row: Harry is to the left of Willie; George is to the left of Harry. Which boy is in the middle?

17. If I have more than a shilling I shall either go by taxi or by train: if it rains I shall either go by train or by 'bus. It is raining, and I have half a crown. How do you think I shall go?

18. On one side of my street the houses all have odd numbers, beginning with the grocer's, which is No. 1. On the other side the numbers are even; No. 2, the baker's, being opposite No. 1. My house is No. 16. Walter is my next-door neighbour: you pass his house as you come up from the

baker's just before you get to mine. What is the number on his door ?

*19. In cold, damp climates, root crops like potatoes and turnips grow best ; in temperate climates there are abundant pastures, and oats and barley flourish ; in sub-tropical climates, wheat, olives and vines flourish ; in tropical climates, date palms and rice flourish. The ancient Greeks lived largely on bread, with oil instead of butter : they had wine to drink and raisins for fruit. Which climate do you think they had ?

Ten Years

20. Some children were asked : " Why are towns nearly always more unhealthy than the country ? " They gave the following replies : (1) " Some country places are by the seaside." (2) " There are more doctors in the towns." (3) " The smoke of the houses and the breath of the people prevent the air from being fresh." (4) " The cottages in the country are dark, tiny, and badly built." (5) " Disease spreads where people are crowded together." Which two children gave the best answers ?

21. " Drinking the sea dry." " Catching the wind in a cabbage net." " Gathering grapes from thistles." " Washing a blackamoor white." " Touching the end of a rainbow." All these sayings mean something that is——? (Give the meanings of all of them in one word.)

*22. The doctor thinks Violet has caught some illness. If she has a rash, it is probably chicken-pox, measles or scarlet fever. If she has been ailing with

a cold or cough she may develop whooping-cough, measles or mumps. She has been sneezing and coughing for some days, and now spots are appearing on her face and arms. What do you think is the matter with Violet?

23. "I sprang to the stirrup and Joris and he:
I galloped, Dirck galloped, we galloped all three."

What was the name of the person referred to as "he" in these lines of poetry?

24. The Duchess of Dustiland's diamonds have been stolen. After the ball at the palace she gave them to her manservant to take home, with instructions to hand them over to her maid at once. When he was half-way home he met a friend, and a few minutes afterwards a man in blue uniform and helmet came up to them and said: "I arrest you for stealing the Duchess of Dustiland's jewels." They were taken to a large building outside which a blue lamp was hanging with the words "Police Station" printed on it. Here another man in uniform took possession of the diamonds, and locked up both the manservant and his friend in a small, bare room for the night. When day dawned, hearing nobody about they climbed out through the window, but could see nothing of either the lamp or policemen. The Duchess is still looking for her jewels. Who do you think is the thief?

*25. There are four roads here. I have come from the South and want to go to Melton. The road to the right leads somewhere else: straight ahead it leads only to a farm. In which direction is Melton—North, South, East or West?

Eleven Years

26. A man was found nearly dead with his throat cut, and on the back of his left arm there was a blood-stained mark of a left hand. The policeman says he tried to kill himself. Do you think the policeman was right?

27. C is West of B : B is West of A. Is A to the North, South, East or West of C?

*28. Father has just come home in a brand new overcoat : there is clay on his boots, and flour on his hat. The only places he can have been to are Northgate, Southgate, Westgate or the City. He has not had time to go to more than one of these. There is no clay anywhere in the streets except where the pavement is up for repair. There are tailor shops only in Southgate, Westgate and the City. There are flour mills only in Northgate, Westgate and the City. I know the roads are not being repaired in the City, though they may be in the other places. Where has father been?

29. The following are some of the occasions on which people shed tears : People laugh till they cry. When they are very unhappy they weep. A fly in the eye makes the tears flow. Peeling onions, scraping horseradish, going through smoke, a cold wind in the face—all make the eyes water. These instances suggest two general causes which produce tears. What are they? Choose your answer from the following phrases : (1) Moderate happiness. (2) Bright colours such as red and green. (3) Germs. (4) Violent emotions. (5) Irritation of the eye-ball. (6) A warm temperature.

30. In our school a third of the school play

football, and a third play cricket. (1) Are there any who play neither football nor cricket? (2) Are there any who play both? (If it is impossible to tell without asking further, say so.)

*31. Where the climate is hot, aloes and rubber will grow: heather and grass will only grow where it is cold. Heather and rubber require plenty of moisture: grass and aloes will grow only in fairly dry regions. Near the river Amazon it is very hot and very damp. Which of the above grows there?

Twelve Years

32. My brother writes: "I have walked over from Byford Wood to-day, where I had the misfortune yesterday to break a limb." Can you guess from this which he probably broke—his right arm, left arm, right leg, or left leg?

33. In the old world the most thickly-populated parts have usually been India, China, and the South and West of Europe. In India and China the rainfall is high in the summer; on the shores of the Mediterranean it is high during the winter; on the shores of the Atlantic it is fairly high all the year round. In the deserts of Russia, Persia and Africa, it is dry all the year round. Africa is very hot; India and China are very warm; South and West Europe rather warm; the deserts of Russia cold. What kind of climate seems to have helped the growth of civilisation most of all—cold and dry, warm and dry, or hot and dry? cold and wet, warm and wet, or hot and wet?

*34. Field-mice devour the honey stored by the humble-bees: the honey which they store is the

chief food of the humble-bees. Near towns there are far more cats than in the open country. Cats kill all kinds of mice. Where, then, do you think there are most humble-bees—near towns or in the open country?

35. My birthday is on December 27, and I am just four days older than Tom. This year Christmas Day comes on a Tuesday. On what day of the week is Tom's birthday?

36. If the train is late he will miss his appointment: if the train is not late he will miss the train. We do not know whether the train was late or not. Can we tell whether he kept his appointment?

*37. I started from the church and walked 100 yards; I turned to the right and walked 50 yards; I turned to the right again and walked 100 yards. How far am I from the church?

Thirteen Years

38. Explain how the following code is worked:

Message (in code) . dpnf up Mpoepo bu podf.

The same (translated) come to London at once.

What is the secret letter for "x" in this code?

39. Dismal Johnny said to Sunny Jim: "If I marry I shall be miserable, because I shall be bothered with looking after my wife; if I don't marry I shall still be miserable, because I shall have no wife to look after me. So in either case I shall be miserable." Sunny Jim replied: "On the contrary, you ought to be happy in either case; for, if you do not marry, you will be happy, because you will not be bothered with looking after your wife, and——" How do you think he finished his argument?

*40. A pound of meat should roast for half-an-hour; two pounds of meat should roast for three-quarters of an hour; three pounds of meat should roast for one hour; eight pounds of meat should roast for two hours and a quarter; nine pounds of meat should roast for two hours and a half. From this can you discover a simple rule by which you can tell from the weight of a joint how long it should roast?

41. I walked 10 yards down High Street; I turned to the left and walked 15 yards down Thomas Street; I turned to the left again and walked 10 yards down James Street; I turned to the left again and walked 15 yards down another street; I turned to the left again and walked 10 yards down that street; I turned to the left again and walked 5 yards. What street was I in?

42. 1 is 1, that is, 1 times 1.
1 and 3 added together are 4, that is, 2 times 2.
1 and 3 and 5 added together are 9, that is, 3 times 3.
1 and 3 and 5 and 7 added together are 16, that is, ——?

Look at the above carefully. Can you see a simple rule for guessing the answers without adding up the figures? Work the following sums yourself; this will help you to find the rule: (i) 1 and 3 and 5 and 7 and 9 added together are——, because this is —— times ——. (ii) What do the first seven odd numbers (1, 3, 5, 7, 9, 11, 13) come to when added together? This is —— times ——. Use the rule to find how much the first hundred odd numbers would come to if added up.

*43. What conclusions can you draw from the following facts? Iron nails will not float in a pool; a cup of pure gold dust weighs nearly twenty times as much as a cup of water of the same size; if you drop a silver sixpence or a copper coin into a puddle, it will sink to the bottom; a cubic inch (about a tablespoonful) of water weighs less than half an ounce; a cubic inch of brass weighs over two ounces; a leaden weight will drop to the bottom of the ocean. Sum up all these observations in one short sentence of the following form: "Most —— are —— —— ——."

FOURTEEN YEARS

44. When you enter my house you will find a window on your right in the side wall of the passage. When the sun sets it shines straight through this window on to the wall opposite. What direction are you facing when you stand in the doorway and look across the street?

45. If the A's have a bigger army than the B's we ought first either to fight the B's, or attack the C's by sea, but not attack the A's; if their army is smaller we should attack the A's first. If the C's have a bigger navy than we, we ought to fight either the B's or the A's, but not the C's. If their navy is smaller, we should first attack the C's by sea. The size of their armies and navies is as follows—

	Men	Ships
A.	7,000,000	300
B.	5,000,000	400
C.	4,000,000	500
Ourselves.	6,000,000	200

Whom should we attack first?

*46. John said: "I heard my clock strike yesterday ten minutes before the first gun fired. I did not count the strokes, but I am sure it struck more than once, and I think it struck an odd number." John was out all the morning, and his clock stopped at five to five the same afternoon. When do you think the first gun fired?

47. Mary has just taken a penny ticket. The trains from this station all stop at Euston, but after that some go to Chalk Farm and Golders Green; others go to Kentish Town and Highgate. They stop nowhere else. The fare to Euston, Chalk Farm or Kentish Town is a penny: to Highgate or Golders Green twopence. Mary did not get in the Golders Green train. To what station do you think she is travelling?

48. They say that in Dodoland hundreds of years ago all the kingfishers had legs about six inches long, and beaks about two inches long, and they used to wade in the water to catch fish for food. The individual birds might differ from one another in the length of their beaks and legs by about half an inch or so—not more. But the offspring of the birds with the shortest legs or beaks would inherit legs and beaks equally short, though again the brothers would differ a little from each other; and similarly with the birds whose parents had longer legs and longer beaks. And the same happened with each succeeding generation. Now in those days the pools were only four inches deep. But they got gradually deeper and deeper; and to-day, where the fish swim, the water is always a foot deep at the very least. Kingfishers of the ancient kind would now-a-days either drown in the deep

water, or starve for lack of food; for they could never learn to swim. What, then, do you think has happened to these wading birds in the course of centuries?

*49. Captain Watts and his son James have been found shot—the father in the chest, and the son in the back. Both clearly died instantaneously. A gun fired close to the person—as, for example, when a man shoots himself—will blacken and even burn the skin or clothes; fired from a greater distance it will leave no such mark. The two bodies were found near the middle of a large hall used as a rifle range. Its floor is covered with damp sand, which shows every footprint distinctly. Inside the room there are two pairs of footprints only. A third man standing just outside the door or window could aim at any part of the room, but the pavement outside would show no footmarks. Under Captain Watts's body was found a gun; no such weapon was found near James. In each case the coat, where the bullet entered, was blackened with gunpowder, and the cloth a little singed. Captain Watts was devoted to his son, and would have died sooner than harm him purposely; hence it is impossible to suppose that he killed him deliberately, even in self-defence. But some think that James secretly disliked his father, and hoped to inherit his fortune at his death. (1) Was Captain Watts's death due to murder, accident or suicide? (2) Was James's death due to murder, accident or suicide?

50. The crust of the earth—that is, the outer layer down to at least fifty miles below the top—consists chiefly of rock and stone. Rock and stone

weigh about three times as much as a bulk of water of the same size. The heaviest materials found in the crust of the earth are metals; but in the outer layer of the earth these are, of course, comparatively rare. The earth as a whole weighs over five times as heavy as a globe of water of the same size. What does this suggest that the interior and middle of the earth are mainly composed of—water, rock and stone, metal or hot gas?

CHAPTER VI

THE MEASUREMENT OF KNOWLEDGE

The heart of the problem may best be made manifest to the reader by supposing that a father brings to him a boy of ten years of age with the request that he should ascertain whether that boy, who is, say, attending a dame's school, is up to the average in his school studies. The father does not want to know anything about the boy's intelligence (no father ever doubts that), but he does want to know whether the school is giving him value for his money—a matter which all parents frequently call into question. Have we any means of telling with any degree of exactness whether the lad is of average proficiency in such rudimentary branches of instruction as reading, composition and arithmetic? Until quite recently we had not. We did not know what the average child of that age could perform: the most we could do was to make a rough guess; and the value of that guess depended purely upon personal judgment, which again depended upon the range and nature of personal experience. We had, in fact, no stable standards of measurement. Each man measured for himself, and each measured with a private yard-stick. And even to-day the position is not much better. We have only just begun to standardise our tests, and have merely arrived at a

few tentative standards of achievement in the simplest of processes. There is an urgent call for the extension of this work.

Many are the motives that urge us to make the attempt. In France the aim has been to find the child who fails to benefit by the ordinary instruction of the *écoles primaires*. Binet's motive in formulating his *Barème d'Instruction* was the same as his motive in establishing his scale of intelligence—the quest of the subnormal child. But other motives are constantly operative. There is a need for standards of comparison between the achievements of children of different races, of different historic periods, of different environments, and under different types and modes of education. Pedagogic records are no less useful than athletic records. It is surely as profitable to know how long it would take a nine-year-old child to add a given column of figures as it is to know how long it would take a two-year-old horse to run a mile. Yet we have records in the one case and not in the other. Finally, there is the need for placing in the hands of the class teacher some means of protection against arbitrary and unreasonable criticism by the headmaster or the inspector.

It is well to be quite clear as to the precise nature of the standards we wish to establish; for there are three distinct possibilities. We may ascertain what children of a given age actually *do* do, or what they *can* do, or what they *ought* to do. The first standard is actual, the second maximal, and the third ideal. And none of them can be either deduced from *à priori* principles or extracted from the inner consciousness of a Board of Examiners.

Our scale must rest on measurable facts, must be built up by a careful investigation into the actual achievement of children under clearly defined conditions. What standard of proficiency in reading, for instance, or in spelling, or in arithmetic, can be expected of London children of a given age under their present conditions of schooling? The simplest and most convenient means of ascertaining this is to devise a suitable test (by no means an easy matter), apply it to as large a number of children as possible, and submit the results to statistical analysis. The scale thus arrived at will afford what may be called, by way of distinguishing it from the other two types of standard, Norms of Performance. These norms have a simple and definite meaning; they can be verified or modified by further testing, and they have a definite range of usefulness.

Is it possible to establish either of the other two standards? And if possible, is it expedient? Is there any point, for instance, in trying to discover what the result in arithmetic would be if all the schools in London were to sacrifice for a time every kind of intellectual activity in order to secure maximal proficiency in this one direction? There evidently is not. Apart from the violence done to the victims of the experiment, the results would be of no value. Can we deduce from the norms of actual performance some sort of ideal standard which would serve as an index of what the children ought to do as distinct from what they do? I myself see no scientific way of doing so. Where ethics and æsthetics have failed, it does not seem likely that pedagogics will succeed. We have therefore to fall back on norms of actual performance as

constituting the only type of standard which proves to be both practicable and useful. Of the other two, one is practicable without being useful, and the other useful without being practicable.

Let us now consider what attempts have actually been made to arrive at standards or norms of instruction. They first appear in the guise of curricula issued by a central authority. In our own country the Board of Education's standards of examination are familiar to all those who remember the days of payment by results. Indeed, they have not yet quite disappeared from the Code of Regulations: they appear in a modified form in the section on certificates of proficiency. It may be said at once that they do not represent any pure type of standard, as the word is used in this book. They are neither actual, maximal, nor ideal. Partaking of some of the worst features of each, they are hybrids of very doubtful pedigree. Based originally on the opinion of what certain authorities at the Education Department thought reasonable, moulded to suit the exigencies of examination, they lacked a solid foundation of objective fact. Moreover, the scheme was used, and, indeed, intended to be used, in such a way as to exemplify what Professor Adams has pointed out to be the one great danger to which all norms or standards are exposed: it was used as a goal, and not merely as a test. Since preparation for annual examinations was at that time virtually the sole aim and purpose of the school, it was inevitable that these standards of examination should become syllabuses of work; should form the ground of the classification of scholars; should, by simple metonymy, give their name to the classes

preparing for the specific examinations ; and, finally, be regarded as norms of performance for children of various ages. Thus the word tried to do several distinct pieces of work, and did none of them well.

As examination syllabuses and schemes of study, the standards were marred by serious faults. In the first place, they were often so vague as to be almost useless. The reading requirement for Standard III, for example, was " To read a passage from a reading-book " ; and for Standard IV, " To read a passage from a reading-book or history of England." Neither sentence tells us more than the first two words alone. No indication is given of the mode of judging whether a child has reached the specified standard, nor any indication as to the amount of progress in reading to be expected of a child between the ages of ten and eleven, the normal ages for " passing " these standards. Secondly, the scale involved too high a standard of mechanical accuracy —a standard which could not be reached without an undue expenditure of time, and a consequent narrowing of the curriculum and loss of interest. With the present wide and generous scheme of studies, so high a level of attainment in one direction would be difficult, if not impossible. Finally, as norms of achievement, the standards of examination were woefully defective. Even supposing that some objective means of testing were prescribed, so that two independent examiners would inevitably arrive at the same results (which was certainly not the case), the standards themselves were arbitrary and lacked the guarantee of a careful scientific study of the normal capacities of children.

Nor did the schedule of standards represent a

real age scale. Although a normal pupil was supposed to pass through each of the seven standards before he reached fourteen years of age, he could only pass one a year; and while a stroke of bad luck would put him back a whole year, no amount of good luck could put him forward a day. The consequence was that retardates were plentiful and accelerates entirely absent.

A very different kind of standardisation is that of Binet. His *Barème d'Instruction* was based on the mean performance of a large number of Parisian children. It supplies, in fact, real norms in three important branches of instruction—reading, number and spelling. But although the aim is laudable, the table published in *Les idées modernes sur les Enfants* is of no great value. As the term "*Barème*" implies, it claims to be a ready reckoner—a rough-and-ready means of finding out whether a child of a given age has profited adequately by his schooling. But its very roughness and readiness constitute its main defect. It is supposed to take only ten minutes to assess a child's scholarship—to fix his position in the age scale; and it is hard to believe that so meagre and hurried an examination can gauge the effect of years of teaching.

Binet's method of judging the reading is based on the position of the pauses made by the reader. If he pauses between the syllables, it is marked "syllabic"; if he pauses between the words, it is marked "hesitant"; and if he does not pause at all, except where the sense of the passage demands it, the reading is marked "current." In all, five grades of proficiency are indicated—subsyllabic, syllabic, hesitant, current and expressive. This

seems a simple scheme, and if it were easy to apply, its obvious lack of delicacy of gradation might be overlooked. But, as a matter of fact, it is impossible to apply it with any feeling of certainty. It is very rare that any child's reading is found to fall entirely within any one of the prescribed grades.

In arithmetic simplification is again aimed at. Binet goes so far as to contend that it is unnecessary to test both addition and subtraction, for a knowledge of the latter implies a knowledge of the former. By parity of reasoning, a test in division is claimed to render a test in multiplication unnecessary. Indeed, three simple questions for each child seem to be considered adequate; but each question must be given in concrete form. To illustrate the kind of question he advocates, I will cite his typical example for children between six and seven years of age: "From 19 apples take away 6 apples: how many remain?"

The spelling test is of no value outside France; for it is a test of grammar as well as of spelling—the two being inseparable in the French language. This peculiarity is rendered obvious by the typical sentence given in the *Barème:* "Les jolies petites filles étudient les plantes qu'elles on remassés hier" —a sentence abounding in pitfalls.

Turning from France to America, we find there the home of standardisation. A tremendous volume of work has been done by Thorndike, Courtis, Ayres, Terman, Munroe, Judd and others with a view to reaching as much exactitude as the nature of the subject will permit. But its usefulness for us is marred by its insularity—by the fact that it is American and not English. Their arithmetic tests

abound in dollars and cents, their problems often refer to customs and transactions peculiar to themselves, and the spelling is Websterian. The next serious defect, however, is that the results are not embodied in an age scale, but in a grade scale. We are told what level the children in Grade IV, say, of a certain American school, did, on an average, attain in arithmetic, spelling and reading. But we are not told the actual age of these children. Nor can we infer it from the grade. We learn from their educational papers that the ages in the grades are not what they are theoretically supposed to be. They vary from school to school, and from city to city. Moreover, it makes a very great difference whether the tests were given at the beginning, the middle, or the end of the school year. Binet's scales are of universal value because they are age scales; and if the American scales are to be used in England they must be translated from grade scales to age scales. They will then cease to be national and become international.

I must emphasise the fact that the tests are merely tests, and that their value diminishes in the proportion that they deflect the general course of study. Although in those rudimentary habits which are the beginnings of school pursuits this danger may be met by making the tests comprehensive, that device is impossible in those higher and more advanced branches of study where no test can cover more than a small fraction of the whole field. We have, in fact, ultimately to rely on the co-operation and good faith of the teacher. And my experience goes to show that such confidence is rarely misplaced.

It has been urged as an objection to the use of norms

that they refer to the most measurable school subjects, and the most measurable are the least valuable. There is thus a danger that the teacher will concentrate his efforts on the more mechanical aspects of school work, to the neglect of the higher and more spiritual aspects. As a matter of fact, this fear is groundless. Norms in the mechanical subjects are just as much a protection as a menace. They will probably show that in some schools the mechanical work is too good : it is so far above the normal that too much time and attention have been devoted to it. When the norm is reached further "drill" is for the time unnecessary. The teacher would realise that he could with a clear conscience reduce the drudgery and render the work more recreational and more broadly cultural. What is really aimed at in establishing the more mechanical norms is the discovery of that degree of automatism which best serves the interests of the higher processes of thought. And in any given school it may be just as necessary to slacken the standard as to string it up.

CHAPTER VII

DISTRIBUTION AND DISPERSION [1]

When we measure a multitude of natural objects belonging to the same genus or class we find that the results tend to assume a determinate shape. The magnitudes occur in accordance with a general law. Let us suppose, for instance, that we start measuring the heights of adult Englishmen by making one after the other stand against a post and marking each height with a horizontal stroke. If we choose our specimens at random we shall soon find our marks tending to crowd round a fixed height. By the time we had measured 100 the marks would appear somewhat as in Fig. 1. The largest number of marks would probably fall between 67 and 68 inches, and the numbers would tail off as they diverged in either direction from this central position. As the number of men increased we should find our range of heights extending, for we are much more likely to find a giant or a dwarf among 10,000 men than among fifty. By the time 8000 had been measured we should probably find our range extending from 57 to 78 inches. Let us now

[1] For a fuller treatment of this subject the reader is referred to Nunn's *Exercises in Algebra*, Vol. II., Sec. IX. (Longmans).

Some excellent diagrams illustrating distribution and dispersion will be found in Burt's *Educational Abilities* (King and Son).

take our 100 men, or, for statistical convenience, let us say ninety-nine, and put them to stand in a row in order of height so that the tallest stand at one end of the row and the shortest at the other. An unreflecting observer would expect the line joining the tops of their heads to be a straight line.

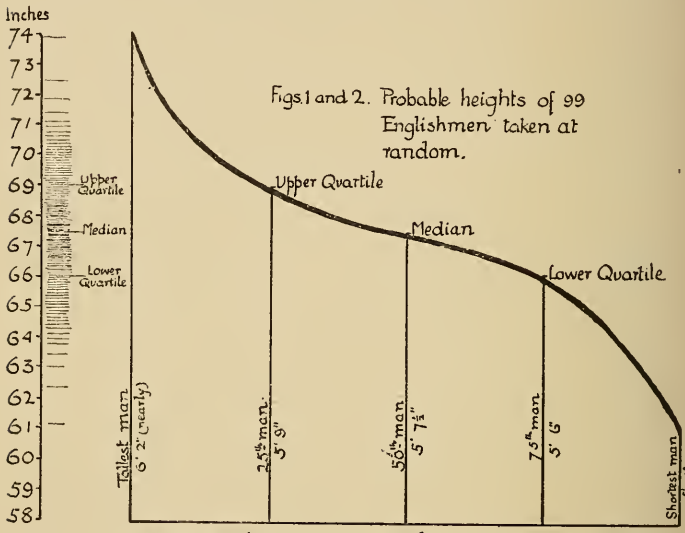


Fig. 1. Fig. 2

In fact, however, the line tends to take the forms of an Ogive—the curve represented by Fig. 2. This, indeed, could be deduced from the facts recorded in Fig. 1. Let us arrange the results in yet another way. If we erect a column, A (Fig. 3), proportional to the number of measures that fall between 67 and 68 inches, place to the right another column, B, proportional to the number of measures

that fall between 68 and 69, and so on for the other groups of measures, we shall get what is called a frequency-column-graph, or a surface of frequencies.

It is clear that we have not been dealing with discrete units like the abstract numbers 1, 2, 3, 4, etc., but with continuous magnitudes. For there is no limit, except those imposed by our instruments and our sense-organs, to the number of

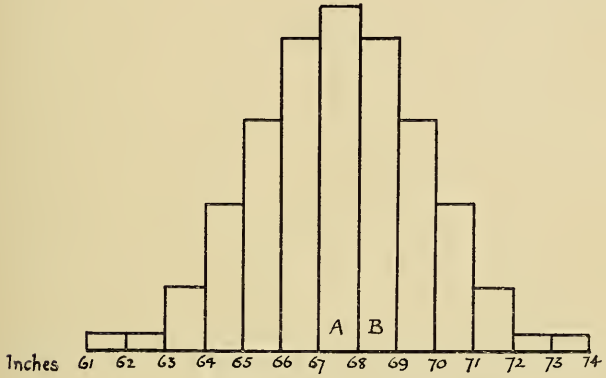


Fig. 3 Frequency-column-graph.

different measurements that may be made between, say, 67 and 68 inches. Nor is there any fixity about the group-range. We have chosen one inch, but we might just as well have chosen half an inch, or a quarter of an inch, or, indeed, any range we liked. It is obvious, however, that the smaller the range of statures the smaller the number of men whose statures fall within that range; and to show the general scheme of distribution it is most convenient to choose one's range in accordance with

the number of men measured. If the numbers are very large we can conveniently make our columns thinner; and by increasing the numbers and diminishing the range we find the zigzag formed by the tops of the columns tending towards a continuous curve, as in Fig. 4. This peculiar bell-shaped curve is known as a frequency curve. The particular type here illustrated, to which stature closely conforms, is called the Curve of Normal Distribution.

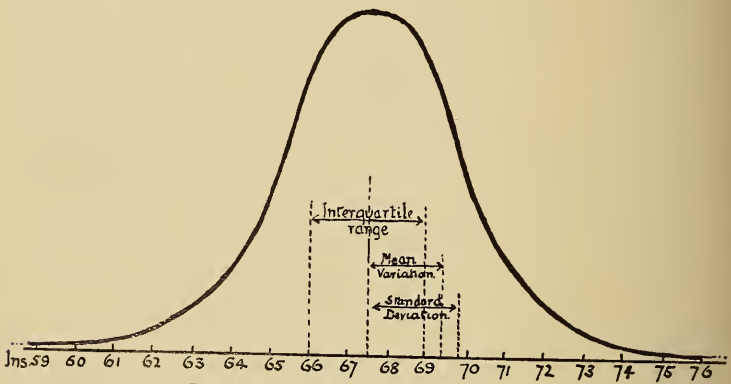


Fig. 4. Curve of normal distribution.

If instead of taking the men's heights we had taken their weights, we should have found the measurements distributed in almost the same way.

Finally, when instead of a physical trait we measure a mental trait we appear usually to get the same type of result. A number of girls, say, of eleven years of age exhibit degrees of educational ability which, when carefully tested, prove to be distributed in much the same way as their heights or weights. In other words, examination marks

should follow the same law as the measurements of other natural features: they should roughly conform to the Curve of Normal Distribution.

This curve is also called the Curve of Probability and the Curve of Error. The term Curve of Probability derives from the fact that the curve shows the frequency of a series of events composed of a number of factors when the presence or absence of a given factor is determined merely by chance. Suppose, for example, I have two coins, toss them up again and again, and make a record of the number of times there is no head, the number of times there is only one head, and the number of times there are two heads. I shall find the score for the three cases tending towards the proportion of 1, 2 and 1. This, indeed, could be deduced by the laws of probability from the fact that there are but four possible ways in which the coins can fall, viz. TT, TH, HT and HH; and out of these four "no head" appears once, one head twice, and two heads once. If instead of the presence of heads we had considered the absence of heads, or the presence or absence of tails, we should have got the same result. The same kind of thing, too, will happen if we toss three coins a large number of times, except that the proportion would now be that of 1, 3, 3 and 1. The case of no head (or three tails) would occur about once every eight throws, one head about three times, two heads about three times, and three heads about once. With four coins the probable ratios would be 1, 4, 6, 4 and 1. The reader will by this time have observed that the proportions I have given are the coefficients of the expansion of $(x + 1)^2$,

$(x + 1)^3$ and $(x + 1)^4$; and if he will proceed to expand the higher powers of the binomial and represent them by frequency-column-graphs, he will find that the graphs gradually approximate the curve in Fig. 4—the curve of probability.

The term Curve of Error is a relic of an older nomenclature. The curve was, in fact, first studied in connection with errors of observation, particularly in astronomy; and a tendency arose to call all divergences from a mean "errors." In gunnery, too, it was natural to regard as errors all the shots that missed the mark.

It should be strongly emphasised that the curve is an ideal curve. It is seldom or never actually reached: it merely marks a limit towards which the values tend—an ideal arrangement to which they approximate more and more as certain conditions (those of pure chance) are fulfilled. Inferences drawn from it, arguments based on it, are never certain: they are only probable. For it is the main characteristic of all inferences from statistical uniformities, that as they increase in particularity they diminish in certainty. They are wholesale truths, not retail truths. The smaller the number of cases the less sure we feel that any general statement will hold good. But the number of children grouped in a class is rarely so small that a central tendency is not observable; and if the results of examining them depart widely from the normal type, it shows that the examination was unsuitable. It was either a bad examination, or a bad sample of children—bad in the sense of not being representative. A picked sample, it is true, always exhibits certain peculiarities; but these

should be allowed for: they should not come as a surprise to the examiner. The results of an examination are, in fact, a standing criticism of the examination: they search the suitability of the questions and the correctness of the marking.

From what we have already said certain important principles may be deduced—

(1) On an ideal examination mark-sheet, where the candidates have been ranked in order of merit, the degrees of difference between consecutive marks are not equal. They are greatest at the extremes, and gradually diminish as they approach the middle.

(2) If equal ranges of ability be taken, the number of cases falling within the central range is large, and the number within the extreme ranges small. If, therefore, we arrange the examinees in three groups, A, B and C, representing equal ranges of ability, the B group should be the largest, and the A and C groups should be about the same size. Similarly, if we classify them in five groups, A, B, C, D and E, the C should be the largest group, the B and D the next, and the A and E the smallest.

(3) A series of tests of any mental quality or attainment should be so devised, and the results so marked, that nobody should gain full marks, nobody gain none, and the average gain about half marks. This is an ideal examination which is never in practice attained. When the aim of the examination is to pick out the brightest intellects these conditions do not hold good.

(4) The range of marks between the median, or middle mark, and the quartile (the mark half-way between the middle and the ends) is less than that between the quartile and the extreme.

It must not be thought that every test the teacher sets his class can give results which accord with these rules. It may be, for instance, that the test is only intended to divide the class into passes and failures, to find out those who can do a certain simple thing and those who cannot. This sheep-and-goat classification is extremely useful for teaching purposes, and may involve any degree of inequality in the sizes of the two groups.

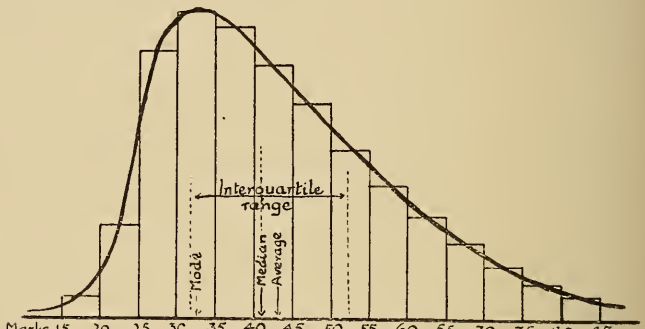


Fig. 5 Skew Curve

Nor must it be assumed that the series is always symmetrical. In fact, the curve is as often as not skewed, as in Fig. 5 : the numbers crowd towards the right or towards the left. There is a general rise and a general fall, but one is often steeper than the other.

The only way to describe a series of measurements fully is to state all the facts, either by giving the actual measurements or by representing them by a graph. But we rarely use all these facts : we are generally content to know their central tendency, and to take that value as representative of

the whole series. When the series is normally distributed the central tendency is represented by the average or arithmetical mean. If the terms of this normally distributed series be arranged in order of magnitude the mean is also the median or middle term. But when the series is not symmetrical the mean is not in the middle, and its right to represent the series has been disputed. It is claimed that the median represents it better. Take, for example, the series (*a*) 1, 2, 3, 3, 3, 4, 19. The average here is 5, but it is clear that the median, 3, is a better index of the central tendency than 5. It may, in fact, be stated generally that when erratic items disturb the tenor of the series, the median is better than the mean as a measure of the central tendency. But there are, on the other hand, other abnormal occasions when the average is the better measure. Take, for instance, the series (*b*) 0, 1, 1, 7, 7, 8, 11. The median, 7, is clearly too high. The average, 5, is better. And in the series (*c*) 0, 0, 0, 8, 9, the median, 0, would be entirely misleading.

It is sometimes suggested that the mode—the most frequent value (*e. g.* 3 in series (*a*), and 0 in series (*c*))—be taken as the representative value. But in some series no value occurs more than once; in others it is impossible to decide between rival claims (the 1 and the 7, for instance, occur with equal frequency in series (*b*)); in others, again, the mode occurs at the beginning or end of the series, as in series (*c*), and is a bad representative.

It may be argued that the series of marks we have been citing violate the very rules we have already laid down: they depart widely from the

normal, and indicate either very bad testing or very bad sampling. The charge of bad sampling may be admitted, for the class teacher cannot always pick his samples: he has to test what children he has. Moreover, he has to measure the success of his teaching—to find out the extent to which he has imparted knowledge or fostered understanding; and here irregular results may reasonably be expected, especially if the numbers are small. I admit, therefore, that for the purposes of emphasis I have taken extreme cases—possible cases rather than probable cases. But these infrequent cases illustrate conditions which are by no means infrequent. They magnify peculiarities that are nearly always present.

When the series is of normal distribution, it does not matter whether we take the mean, the median or the mode; for they all three coincide. When, however, the distribution is not normal we have to choose between them. The mode we may dismiss at once as never quite suitable, and often quite inapplicable. Of the other two sometimes one is more suitable, sometimes the other; and rarely is either a really bad representative value. For simplicity of calculation the median is undoubtedly the better.

When we have learnt the average or central tendency of a series we have learnt the most significant thing about it. But there are other things we want to know; and the most important of these other things is the degree of scatter—the closeness or looseness with which the various values cluster round the average. In series (*d*) 4, 5, 5, 6, 6, 6, 7, 7, 8, for instance, there is very little scatter; no value

deviates from either the mean or the median by more than 2. But in the series (*e*) 1, 1, 2, 3, 3, 5, 8, 14, 17, which has the same average as (*d*), the deviations are very much greater. To evaluate these deviations the obvious plan is to treat them as we treat the original series, that is, find their average. In series (*d*) the deviations are respectively 2, 1, 1, 0, 0, 0, 1, 1, 2, and their average is ·8. This is their *mean deviation from the mean*, or, more briefly, their *mean deviation* or *mean variation*. The mean deviation of series (*e*) is 4·6, that is, more than five times as great as that of the first series.

In the same way we can calculate the *mean deviation from the median*, another index of dispersion.

The simplest and quickest way, however, to measure the general degree of deviation is to extend the principle of the median, and after finding the median, which divides the series into halves, to find the median of each of the halves. In other words, we arrange the series in order of magnitude, and divide it into four equal parts. The value one quarter of the way up is the *lower quartile*, the value half the way up is the *median*, and the value three-quarters of the way up is the *upper quartile*. The difference in value between the upper and lower quartile is known as the *interquartile range*. Since this range is an index of dispersion both above and below the median, it is usual to halve it, and thus get the *semi-interquartile range*, or *quartile deviation*. This is, in fact, a convenient way of averaging the deviations of the two quartiles from the median. In the series (*f*) 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 10, for instance, 3 is the lower

quartile, 5 the median, 8 the upper quartile, 5 the interquartile range, and $2\frac{1}{2}$ the quartile deviation. Sometimes the lower quartile is called the 25 percentile, the median the 50 percentile, and the upper quartile the 75 percentile. The semi-interquartile range is practically the same as the *probable error*. It is so-called on what seems to be a *lucus a non lucendo* principle; for it is neither an error, nor is it probable. Error means deviation, as in the phrase "curve of error"; and it is only a probable deviation in the sense that since twice its amount includes half the number of values, any particular value is just as likely to fall without that amount as within it.

We have one other means of measuring the degree of variation from the central tendency—*standard deviation*, or root-mean-square-deviation. Instead of taking the mean of the individual variations from the mean, we take the mean of their squares and then find its square root. This takes a longer time to calculate than any of the others, but it has certain advantages. In calculating it we have not to commit the algebraical solecism of adding terms of opposite signs as if they were signless, as in estimating the other measures of deviation; for the squares of all numbers are necessarily positive. Moreover, by squaring, extreme deviations are magnified, and given their due weight as indications of individual ability. The very items that cause us trouble when we try to find the central tendency are of supreme service in calculating the correlation of abilities. The best-known method of finding the correlation coefficient involves the finding of the standard deviation. Those, therefore, who wish to make further calculations from an array of values

are prone to use the standard deviation as a measure of variation instead of the simpler measures, or of the mental age. Mr. Burt does so. He suggests that for strictly scientific purposes it should be used in the schools as a unit of measurement.

In a normally distributed series the quartile deviation is less than the mean deviation, and the mean deviation less than the standard deviation—a fact illustrated graphically in Figs. 4 and 5.

I have described various means of finding the central tendency of a list of examination marks, and various means of finding the deviation from that central tendency. From among these there is in actual use in the schools only one way of recording the central tendency, and none of recording the deviation. The average, or arithmetical mean, is the only statistical device in common use. The result of a test in arithmetic is supposed to be adequately reported and recorded when the average number of sums right is given. And as I have already shown, the way of the average is quite a good way: it is at least as good as any other. But it is not enough. Some mode of expressing the degree of dispersion is also necessary—the mean deviation, for instance. As a matter of simplicity in working, however, and completeness of statement, I would commend the method of the median and quartile deviation. To record a result it is only necessary to give five values—the lowest, the lower quartile, the median, the upper quartile and the highest. This will give more information in a small compass than any other method. The mean deviation affords no clue to the symmetry of the series, for it does not differentiate the deviations below

the average from the deviations above the average. But the quartiles reveal to us at a glance both how and how much the several values are dispersed (see Fig. 5). In fact, we learn from the five items mentioned above the range of the series, its central tendency, its degree of dispersion, and its mode of dispersion.

The teacher unfamiliar with statistical terminology will no doubt consider these distinctions irritating and useless. Irritating perhaps they are, but not useless. For different investigators affect different systems. Some find it more convenient to reject the average and use the median, others cling to the older scheme. Those who wish to base elaborate calculations on their data prefer to measure dispersion by the standard deviation; while those who desire quick practical results prefer the simpler measures, the mean variation or the quartile deviation. The few explanations I have given should enable the reader to adjust his mind to these varying usages.

I will try to illustrate some of the ways in which a broad knowledge of the principles of distribution may be useful to the teacher. Let us suppose that he has to mark forty composition exercises written by the boys in his class. It is doubtful whether he can arrange them with any degree of confidence into more than ten groups. Although a teacher sometimes tries to mark on a scale of 100, awarding one boy, say, 73 marks and another 74, it is idle to pretend that he can make such fine discriminations. Generally speaking, ten grades are as many as he can manage; and on a ten-grade basis it is clear that the forty papers in question cannot all receive

different marks. To examine his distribution the teacher should, after marking, draw up a frequency graph as in Figs. 3, 5, 6 and 7. He will then see at once whether he has judiciously apportioned his scores. If the bulk of the marks are somewhere near the middle it is presumptive evidence of careful marking. As an example of bad marking I will take an instance that occurred some years ago in the top standard of an elementary girls' school. At the terminal examination the thirty-five girls in the class were examined by the head mistress in composition, 20 being the maximum marks allowed. Twelve girls got full marks, five got 19, twelve got 18, and six got 15. If the reader will plot this result either as an ogive or as a frequency graph, he will note that the principles I have tried to expound are flagrantly violated. In point of fact the twelve papers that were awarded full marks were of very unequal merit, and it is doubtful whether the best of them deserved more than 14 marks out of 20.

Both composition and reading are quite frequently marked in this haphazard fashion. If more than one pupil receive full marks it generally means that the best of these pupils is getting less than justice. Towards the middle of the scale "ties" are in the natural order of things, but the further from the middle we get the more unlikely is equality to be found. A test, therefore, that fails to differentiate, and spread out the pupils at the higher and lower ends of the list, is either bad in itself or badly marked.

Let us now consider a subject in which it is possible to calibrate more finely. Mathematics, for

instance. Here, with a sufficiently searching examination, it is quite possible to find 100 distinct grades of proficiency. If the base line of a frequency graph be spaced out in equal ranges of from 0 to 5 per cent., from 5 to 10 per cent., from 10 to 15 per cent., and so forth, and columns raised proportional to the number of pupils whose marks fall within these ranges, the result will be a graph or curve of which Figs. 5 and 6 are possible varieties. The skew shown in Fig. 5 may be due to an imperfect grading in

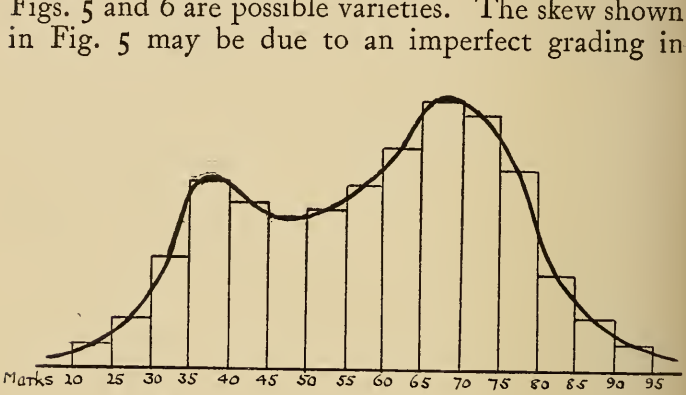
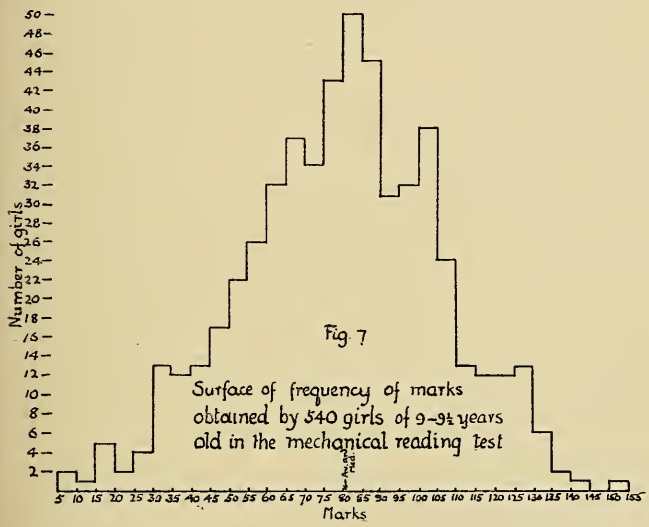


Fig. 6 Curve with two crests

the difficulty of the questions, or to some peculiarity in the class itself. When there are two distinct crests, as in Fig. 6—and especially if this duality reappears in other examinations—it points to a serious lack of homogeneity in the class. These are really two central tendencies due to the presence of two natural groups. We get this kind of curve when we plot the heights of a mixed group of men and women. One crest will indicate the central tendency of the men, and the other the central tendency of the women. A double summit that

sometimes appears in the plotted results of a competitive examination probably indicates two distinct types of candidates—those who work for the examination and those who do not. Peculiarities in the curve of distribution, although they fall short of proof, always suggest points for investigation.


Number of girls
Marks
Fig. 7
Surface of frequency of marks obtained by 540 girls of 9–9½ years old in the mechanical reading test

I would particularly warn the reader not to expect a smooth curve from so small a number of cases as a single class provides. Rather should he expect something like what is seen in Fig. 7, where I have plotted some of the results obtained in the reading test described in Chapter VIII. Although there are 540 cases dealt with, yet there is much irregularity in the rise and fall of the frequencies. Indeed, so conspicuous is the fall at 90–95, or rather the rise at 100–105, that an explanation

seems to be called for. As, however, these features do not reappear in the corresponding frequency surface for boys, I conclude that the causes are accidental, and not due to any peculiarity in the test.

Another way of using the notion of normal distribution in ordinary class work may be indicated. Suppose the teacher has to mark forty drawings. Their absolute merit is beyond his, or indeed anybody's, power to determine: his concern is to appraise them relatively to one another. He will find it a good plan to spread the drawings out on a large table, and proceed thus: If there is one copy which is manifestly the best, mark it 10; and if there is a copy which is manifestly the worst, mark it 1. Then look for the second best—there may be two or more. Mark each of these 9. Then look for the next to the worst, and mark these 2. Work thus towards the middle, making the groups larger and larger as the middle is approached. Here we have a working conception which will be found serviceable as a guiding principle, but dangerous as a hard-and-fast rule. It is, indeed, very improbable that a set of papers if rightly marked will tamely fall in with a preconceived scheme, for each set will have peculiarities of its own which must be recognised; but it is equally improbable that the drawings will depart widely from the general scheme I have outlined above. In the final issue the marks should fit, not the scheme, but the papers.

In Chapters VIII and X, I give two standardised tests in reading and seven in arithmetic. The norms appended represent the central tendency (the average

or the median) of the results obtained by applying the tests to several thousands of children in a large variety of schools. They indicate the normal achievement at the exact age mentioned; and the way to use them is simple. If, for instance, a teacher gives the silent reading test to his pupils, whose average age is ten years three months, he should expect an average score of 11; for the table of norms gives 10 for ten years of age, and 14 for eleven years of age, and his boys are a quarter of the way between ten and eleven.

Judging from the score made in certain individual schools, the mean variation for the written arithmetic tests is on the whole nearly 50 per cent. of the average reached. If, for instance, the norm is 6, the mean variation is nearly 3. It is larger even than this for the younger children, but considerably smaller for the older. It should be borne in mind, however, that in collating my results it was not children of the same class or grade or standard that were grouped together, but children of the same age. And unless the degree of dispersion in the class is less than the degree of dispersion in the age group, the school is no better organised (for that particular subject at least) than if the classification of scholars were based strictly upon age. The smaller the mean deviation the more homogeneous is the class.

CHAPTER VIII

READING

A.—As a Mechanical Art

What is the fundamental and essential factor in the ability to read? Binet seems to assume it to be fluency; for the scale for marking reading in his *Barème d'Instruction* is based upon the frequency of such pauses as are not necessary to elucidate the meaning of the passage read. The examinees are classified according as they pause after the letters, the syllables, the words or the phrases. More recent investigators, discriminating between the silent reading we do for our own use and pleasure and the oral reading we do for the benefit of others, claim the former to be the more important, and select the comprehension of the material read as the first essential, and speed of reading as the second. Compared with these, intonation, expression and pauses are regarded as unimportant. For this point of view there is much to be said. It is significant of that happy change which is taking place in our schools—the change from reading as an elocutionary display (a very indifferent one at best) to reading as a pleasurable pursuit. But comprehension is a difficult thing to measure; and it presupposes a more rudimentary ability which is quite easy to measure—the ability to translate certain

visual symbols into sounds, whether those sounds are actually produced or are merely imagined. This is the basal ability—the *sine qua non* of reading—the step which no amount of intelligence would enable the learner to dispense with. His intelligence will, of course, eke it out; it will enable him to anticipate coming words, to respond to slighter cues; but ultimately his power to read is rooted in his ability to translate symbols into sounds, and this fundamental ability is what I here set out to measure. It is not the whole of reading; but it is the basal and indispensable part, and it is the part which best lends itself to exact measurement.

The test I adopted after some preliminary experiments is given on the following page :—

The pupil was given a paper on which the test was printed as shown, and was asked to read as fast and as carefully as he could until he was told to stop. The number of words correctly read in a minute—that is, the total number read minus the number misread—gave the score. If the examinee hesitated for more than five seconds over a word and did not seem to be on the point of saying it, he was prompted and told to pass on, that particular word counting as an error. He was thus prevented from being too heavily penalised by his inability to recognise one or two particular words.

It is quite clear that the test does not take intelligence into account; it is designed to measure the bare mechanical art of reading—the degree of facility with which a pupil can translate the symbols of the simplest and commonest words of the mother tongue into the words themselves. Sense material was discarded as tending to confuse the issue. If

One Minute Reading Test

is me on at by so us an it or be to as he of in go up am if no we my ox do the and for but him are can she dog let you not was out try see mix cat now boy saw bit met top run man pet lot get did van bad red cup bee lit pin had ran pen nut big old yet rob gun leg fun lip new fog has sit sly wig mud box ink sat end cut pay fed who six lad wet dry cow his peg tin say eat any far set bud kid pup fox ask egg cab ill use jam all pit got sad tea sky one yes fur act toe her our ten arm rock gone feel that rich till long flat this part foot made upon came mile back sand time said then wall into were done walk much loss seem went with come

an intelligent lad, for instance, starts reading a fairy tale which begins "Once upon a time," the recognition of the first word brings inference into play, and the rest of the phrase is largely a matter of memory. But no such intelligent anticipation is possible in the test used; each word, standing isolated in meaning, has to be read without any help from the context. It has to be read, not inferred.

Much scorn has been levelled at this type of reading. It has even been denied that it is reading at all, and has been contemptuously called "barking at print." But my investigation tends to show that this mechanical response to the seen word, this barking at print—call it what you will—is not only the basis of all reading, good as well as bad, but is also a trustworthy criterion of the interest with which reading is followed as a pursuit.

As a test common to children of different schools and different ages, it possesses many points of advantage. Familiarity with subject-matter, which must always be a variable and disturbing factor where sense material is used, does not enter into the question; there is no subject-matter. A peculiar knowledge of uncommon words avails nothing: there are no uncommon words. However young a child may be, however rudimentary his knowledge, if he can read at all, he can read some of the words in the first three lines; for all the common two-letter words in the language are to be found there. And however proficient a child may be he will find the task of reading the whole of the 158 words in one minute quite as much as he can accomplish.

But these *a priori* reasons in favour of the test

do not dispense with the necessity of testing the test—of comparing, for instance, the results obtained by using this test with those obtained by using sense material, such as an "unseen" passage in the ordinary school reader. Such a comparison I carefully made with ten boys of about eight years of age. The mode of scoring was in each case the same—the number of words correctly read in one minute. The first point to be noted is that although 32 per cent. more words were read per minute in the continuous prose than in my test, the order of merit was, with one trifling exception, the same in both cases. To test one kind of reading is virtually to test the other. The second point to be noted is that the discrete material gives a higher degree of reliability or steadiness; for when the children were put to read the same passages again after five minutes interval, they improved but 7 per cent. with the discrete words as compared with an improvement of 22 per cent. with the continuous prose. The 7 per cent. improvement was perhaps entirely due to a loss of nervousness; but most of the improvement in the other case was doubtless due to the acquired familiarity with the meaning. As a test, therefore, the discrete words are better suited for repeated use.

In estimating the speed of reading, it is well to be clear what precisely we are measuring. For there are four possibilities. We may measure the normal or the maximal speed of either oral or silent reading. The speed measured in this test is the maximal for oral reading.

The question may be raised: how far is the recorded speed a real index of the speed of reading

as distinct from speed of articulation? There is little doubt that adults can read silently very much faster than they can read aloud—very much faster than they can articulate the words. But this is not true of a young child. I found by testing a number of children of 7 years of age that they could speak or recite at the rate of 170 words per minute while they read the test words at the rate of 40 per minute. Again, girls of $9\frac{1}{2}$ years old who read 80 words per minute were able to recite 220 words per minute.

The following norms were obtained by applying the test to the children of 49 schools :—

Age.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.	14 yrs.
Boy's Score.	13	33	53	72	85	115
Girl's Score.	15	38	58	76	88	122

The original experiment stopped at 10 years of age; for I did not regard the test as suitable for older children, with whom the mastery of the mechanical factor may as a rule be taken for granted; but as an afterthought I extended it so as to include children of 14. The results show a considerable flattening of the curve after 10.

Among the 49 schools I included 10 which were reputed to be attended by the poorest children in London, and 9 situated in good residential areas. The results show that both boys and girls in good neighbourhoods are about 6 months (10 marks) in advance of the average; and in poor neighbourhoods the boys are 3 months behind, and the girls 6 months. Thus for the extreme types of home we find a difference of 9 months for boys

and 12 months for girls. A girl of 8 in Dulwich can read as well as a girl of 9 in Bermondsey.

Through the courtesy of a friend, I was enabled to secure results from 6 elementary schools in Lancashire. In general tendency they tallied with the London results; but the Lancashire boys were 3 months ahead of London boys, and the girls 4 months ahead. The 6 schools, however, included no poor schools, and there is reason to believe that the social conditions were above the average.

There is nothing to indicate whether the superiority of the children from good homes is due to heredity or environment—to higher native ability or to better opportunities. Both probably contribute.

The statistics obtained lend support to the popular belief that girls read better than boys. Generally speaking, they are 5 marks or 3 months ahead. But it is a curious fact that this difference does not obtain in slum neighbourhoods. There the boys read quite as well as the girls. This is not merely true of the ten schools taken in a lump; but, with one exception, of each of the 10 schools individually. Whatever the reason may be, whether it is due to the girls being more fully employed minding the baby—there generally is a baby—or doing domestic work, or whether this intrinsic sex difference does not hold good generally, the fact itself seems to admit of little doubt.

Occasionally a reader rushed through the test with little regard to accuracy, and thus made a fairly high score in spite of many blunders; whilst another reader would proceed more cautiously and secure a lower score although he made no

blunders at all. The majority, however, maintained a steady level of cautiousness; and on an average about three mistakes were made by children of all ages. This means that accuracy improved with age, and was roughly proportional to speed; for to get three mistakes out of 88 words read (see score for girls of 10) was nearly six times as creditable as to get three mistakes out of 15 words read (score for girls of 6).

The method of teaching reading in all the schools tested was some form of the phonic. The alphabetical method pure and simple may be said to be obsolete in London, although the children are generally taught the conventional names of the letters either at the same time as their phonic values or after the art of reading has been at least partly acquired. The look-and-say mode of reading into which all other forms ultimately merge, is to a greater or lesser degree incorporated in the other systems. In fact, the system in vogue is mixed, with the phonic element predominating. There are, however, various forms of the phonic method current, some simple and some complex. Perhaps the most systematic and complex of all is the Dale method; and in 19 out of the 49 schools this method is followed. Do the scores at these 19 schools afford evidence of the superiority of this system? How do they compare with the scores at the 30 non-Dale schools? Taken as a whole the Dale boys are 1 month behind when they are under 7 years of age, and 2 months in advance after 7; while the girls are 2 months behind when under 7, and 2 months ahead after 7. But 5 of the 19 Dale schools are included in the 9 schools above referred

to as situated in the wealthier neighbourhoods. And if the influence of home and district be eliminated it is doubtful whether the Dale results are as good as the non-Dale. It is certain that they are no better. Here is a system of high repute, regarded indeed by some as the mark of up-to-dateness in infant teaching—a system requiring special training on the part of the teacher, and involving the use of special apparatus and special books; but judging from the results achieved, it is no better as a means of teaching reading than the ordinary phonic system in common use. It is true that it has other merits. It includes other things besides reading proper, such as printing, drawing, and phonetic analysis—things in themselves valuable; but taken purely from the reading point of view, it fails to substantiate the high claim generally made on its behalf.

What then is the salient characteristic of method in the schools where the test shows reading to be exceptionally good? It is this. In the good schools reading is encouraged as a pursuit, and not merely taught in a series of lessons. The children read for the sake of the story, not for the sake of reading. It is a pleasurable occupation, and not a tiresome phonetic drill, nor yet an elocutionary display. Phonetic drill and elocution have their place in school routine, but the root of proficient reading grows in other soil. In the school where the reading is best of all, even children of six read small fairy tale books for pleasure.

There is one arresting exception to the above generalisation. In a certain infants' school where no private reading takes place, but a large amount of word-building is taken on the blackboard, the

result reaches high-water mark; but in the senior school the curve sinks to the normal: the proficiency is not maintained. It will thus be seen that the same apparent results may be obtained in the earlier period by opposed methods; but while the momentum is kept up by the one method, it is lost by the other. As ever, it is a question of interest.

It should be clearly understood that this test, consisting as it does of words of no connected meaning, is intended to be used as a test only, and never as material for practice, nor even as suggesting the type of material for practice; for I strongly hold the view that in selecting reading-books for children, subject matter is a consideration of supreme importance. If the subject is of no interest to the child, if it would not grip his attention when read *to* him instead of *by* him, then we must regard it to be of the wrong sort; unless, indeed, we reject the view that the main aim and purpose of the teacher of reading should be to hasten the coming of the time when the child will spontaneously take up a book and read it for the sake of what it has to tell him.

Although the test is oral, it must not be inferred that practice should be entirely oral, or indeed mainly oral. For the out-of-school reading, for which the school reading is a preparation and a training, is almost entirely silent: it is reading as a device for getting ideas. And the training of any specific activity should always, as far as possible, take the form of that activity. It is not always possible. At the earliest stage of reading, for instance, much drudgery in associating symbols with sounds is necessary before the process is sufficiently mechan-

ised for the sense of what is read to stand out clearly and predominantly in the mind. Indeed, for some time, the child, unless he reads aloud—unless there is somebody for whom he feels he is doing something—will not read at all. But the sooner this stage is passed the better.

Again, fluent reading, regarded merely from the mechanical point of view, is mainly a matter of practice, and the amount of practice given in school, especially where individual class reading—reading in turns—is adopted, is so small that many years pass before a fair degree of fluency is attained.

The reasons are now clear why I regard this test as useful for the early stages of reading only. As the child grows older the aspect of reading which becomes increasingly important is the extent to which it can be employed as a thought-getting device—the accuracy and rapidity with which the child absorbs the meaning of what he reads privately.

It has often been pointed out that to read mechanically is one thing; to understand what is read, another. And although it will be granted that a child may do the former without the latter, it is equally certain that he can never do the latter without the former. Until, indeed, a child can read about 100 of the test words per minute, the mechanical art of reading cannot be said to have been completely mastered; and the utilisation of the test is profitable as a means of measuring his fluency; which is roughly a measure of the practice he has had; which, again, is roughly a measure of the interest he takes in reading. Beyond that stage other tests overshadow the mechanical one. Reading, indeed, cannot in any sense be

regarded as a simple process, nor can the same means be adopted for measuring proficiency in the earlier and later stages of acquiring the art.

B.—As a Means of Acquiring Ideas

At the age of nine most children have acquired a fair degree of facility in the mechanical art of reading. Henceforth reading is to them either an elocutionary exercise (reading aloud), or a means of getting ideas (silent reading). And of these two the more important is the latter. It is no exaggeration to say that in adult life ninety-nine out of every hundred books are read silently. The essential aim, therefore, in the teaching of reading, should be to give the pupil the power to absorb meaning from the printed page. If he is to gain either pleasure or profit he must understand what he reads. And the measure of his progress in understanding is virtually the measure of his progress in reading. But how are we to measure understanding? It is clear that the time-element cannot be ignored. What we have to determine is the rate at which the pupil can master the meaning of a given piece of prose or poetry.

There are in use, in the United States, at least eight different types of silent reading tests. They generally consist in allowing the pupil a fixed time to read a given passage, and then ascertaining how much of the meaning can be reproduced. But here we encounter the difficulty that makes composition so hard to assess. How can we weigh ideas? How can we measure meaning? Daniel Starch, in his silent reading tests, tries to get over

the difficulty by getting the pupil to write out what he remembers of the passage read, and by counting the number of words written which correctly express the thought. This is equivalent to marking a piece of composition by its length. Some of the other systems require a complicated key by which the reproductions may be scored. The scheme, however, which seems to be most widely used in America, is the Kansas Silent Reading Scale. It comprises a series of sixteen exercises which carry marks proportional to their difficulty. Exercise 6, for instance, which is valued at 2·3, runs as follows: "In going to school James has to pass John's house, but does not pass Frank's. If Harry goes to school with James, whose house will Harry pass, John's or Frank's? . . ." The examinee has to fill in the blank at the end. This question is typical of the series—a series which admittedly tests reasoning: but does it test reading? The pupil's mind is required not merely to follow the meaning of the sentence, but to go beyond it; and it is quite conceivable that a child may be able to follow a plain narrative with ease and rapidity, and yet be very slow at dealing with puzzles of the Kansas kind. In any case the exercises differ *in toto* from the kind of matter people generally read. The last book one thinks of reading is a book of conundrums.

Another peculiarity common to the American silent reading tests, is that more than one series is used. There is generally one for the lower grades, one for the middle grades, and one for the higher grades. And although the several series are supposed to be so adjusted in the matter of difficulty, that

the norms for the various school grades increase regularly, the adjustment is in point of fact never perfect. In the Kansas tests, for instance, the norms for Grade V and Grade VI are nearly the same.

There is, further, the outstanding disadvantage to the Englishman, that the tests and results are arranged by grades and not by ages. The scale is never, except by precarious inference, an age scale.

The test I have devised for my own use is as follows :—

Silent Reading Test

(3 *Minutes*)

One fine morning in spring a robin flew down to the brink of a stream to quench his thirst. Seeing a trout in the water he began to talk to him. " I have often wondered," said he, " how you manage to keep alive. If I tried to stay under water like you I should be dead in a few minutes. And even supposing I could remain alive I should feel miserably cold in the chilly water without either fur or feathers. Please tell me why you are not drowned, and why you do not perish with cold." " You should never ask two questions at once," said the trout. " Quite right ! " croaked an old crow who had heard their conversation and had alighted on the bank beside the robin. He was very old and very wise and very inquisitive. Some thought that it was because he was inquisitive that he was so wise, others thought that his wisdom came from age and experience. Certain it was that he was regarded as the most learned of all the birds of his time, and that he used such long words that the little birds, beasts and fishes could rarely understand what he was

talking about. To return to the robin's questions, the trout replied: "Why don't I get drowned? Why, one does not drown in the water: one drowns in the air." "Nonsense!" said the robin. But the crow looked so severely at him that he trembled, and drooped his tail by way of apology. "As for feeling cold," continued the trout, "I don't know what you mean." The robin was about to say "Liar!" when he caught the crow's eye and restrained himself. Then the crow, having held up his right foot to enjoin silence and attention, delivered himself thus: "It was held by Aristotle, and his opinion is confirmed by modern scientific theory, that there is no living creature that can exist in any and every environment: each requires surroundings suitable to its bodily structure and its bodily functions. To secure a supply of oxygen, which is necessary to maintain the purity and temperature of the blood, each animal is provided with organs which are adapted for extracting this element from the atmosphere, where it is present in great abundance. In land animals lungs serve that purpose; in fishes, gills. Gills are so constructed that they can only take up the oxygen that is dissolved in water. In the air the gills adhere together and the poor fish dies of asphyxiation. Our friend the trout was therefore quite justified in asserting that the atmosphere drowns fishes. As for his avowed ignorance of the distinction between heat and cold, we must not rashly accuse him of falsehood. It is a well-known psychological fact that sensitiveness to heat and cold is dependent on certain specialised nerve endings in the skin." He paused here to see the effect of his oration on his audience.

While speaking he had been gradually closing his eyes in order to think better; but at this point he opened them wide and found to his disgust that the trout had disappeared and that the robin was struggling with a big fat worm about twenty yards away.

A double sheet, with the story printed inside, is distributed, and the children told that they will be given three minutes to read it, and that they will afterwards be tested to see how much they can remember. On the word of command the children must open the papers and read silently. At the end of three minutes the papers are collected and the following completion test given out to each child, with instructions to supply the missing words except those marked (x).

Completion Test
(Unlimited Time)

One fine morning in (1) a (2) flew down to the brink of a (3) to quench his thirst. Seeing a (4) in the water he began to talk to him. "I have often wondered," said he, "how you manage to (5) (6). If I tried to stay under water like you I should be dead in a few minutes. And even supposing I could remain (x) I should feel miserably (7) in the (x) water without either (8) or (9). Please tell me why you are not (10), and why you do not perish with (x)." "You should never ask (11) questions at once," said the (x). "Quite right!" (12) an old (13) who had heard their conversation and had alighted on the bank beside the (x). He was very (14) and very (15) and

very (16.) Some thought that it was because he was (x) that he was so (x); others thought that his (17) came from (18) and (19). Certain it was that he was regarded as the most (20) of all the birds of his time, and that he used such (21) (22) that the little birds, beasts and fishes could (23) understand what he was talking about. To return to the (x) questions, the (x) replied: "Why don't I get drowned? Why, one does not drown in the (24): one drowns in the (25)." "(26)" said the (x). But the (x) looked so (27) at him that he (28), and drooped his tail by way of (29). "As for feeling (x)," continued the (x), "I don't know what you mean." The (x) was about to say "(30)," when he caught the (x) eye and (31) himself. Then the (x), having held up his right foot to enjoin (32) and (33), delivered himself thus: "It was held by (34), and his opinion is (35) by (36) scientific theory, that there is no living creature that can exist in any and every (37): each requires (38) suitable to its bodily (39) and its bodily (40). To secure a supply of (41), which is necessary to maintain the (42) and (43) of the (44), each animal is provided with (45) which are adapted for extracting this (46) from the (47), where it is present in great (48). In land animals (49) serve that purpose; in fishes, (50). (x) are so constructed that they can only take up the (x) that is (51) in water. In the air the (x) adhere together and the poor fish dies of (52). Our friend the trout was therefore quite (53) in asserting that the atmosphere (54) fishes. As for his avowed (55) of the distinction between (56) and (57), we must not

rashly accuse him of (58). It is a well-known (59) fact that (60) to (x) and (x) is dependent on certain (61) nerve endings in the skin." He paused here to see the (62) of his (63) on his (64). While speaking he had been gradually closing his eyes in order to think better ; but at this point he opened them wide and found to his (65) that the (x) had (66) and that the (x) was struggling with a big fat worm about (67) yards away.

The most convenient way to administer the completion test, is to distribute strips of paper upon which the missing words are to be written. If the lines on the paper are numbered from 1 to 67 it will facilitate marking. One mark is awarded for each word correctly given. As it is a test of substance memory, and not rote memory, synonymous words are accepted. General terms, however, must not be regarded as substitutes for particular terms.

The following synonyms and variations are permissible : 3, brook, river ; 5, remain ; 5 and 6, exist ; 8 and 9, in either order ; 12, said ; 14, 15 and 16, in any order ; 18 and 19 in either order ; 21, big ; 23, hardly, scarcely ; 25, atmosphere ; 27, sharply, crossly ; 31 checked, stopped ; 35, upheld ; 38, conditions ; 39 and 40, in either order ; 42 and 43, in either order ; 43, heat ; 46, oxygen ; 47, air ; 48, quantities ; 52, asphyxia, suffocation ; 53, right, correct ; 56 and 57, in either order ; 58, lying, untruth ; 61, special, particular ; 63, speech, words ; 64, listeners, hearers ; 66, gone, vanished. In every other case the exact word must be given.

This test has the merit of being simple and workable, of being applicable to all readers from the ages of nine to ninety, and of being objective in the

sense that two independent examiners will, if they follow the instructions, inevitably give the same mark for the same achievement. The examiner may, if he wishes, examine himself. It will be seen that the narrative is easy to begin with, and gradually increases in difficulty, so that the dullest reader will be able to remember some, and the brightest fail to remember all. I have tried as far as possible to select, as missing words, those which indicate that the meaning of the text has been understood, and yet are not such as an intelligent pupil would infer from the context. The first blank, for instance, cannot be filled by reasoning, for there are many possibilities. The missing word may be "Summer," or "April," or any other season or month. And the second word may be any one of the numerous species of birds or insects. In fact, the missing words are not inevitable words: they really test whether the pupil has read the piece, has understood it, and has remembered it.

To prevent the examinee from inferring the missing words at the beginning of the piece from what is said later on, I have omitted the words that afford a ground for retrospective inference and substituted (x). If, for instance, I had printed "trout" instead of (x) in the twelfth line, a shrewd lad would at once know that the fourth missing word was "trout." In order to ascertain experimentally, to what extent inference could supply the place of direct knowledge acquired from actually reading the piece, I submitted the second paper to a number of intelligent children who had not read the first. In no instance were more than three words guessed correctly.

By applying the test to about 5000 children from nine to fifteen years of age, I arrived at the following norms :—

Age.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.
Score.	6	10	14	17	19	21

Through the courtesy of Miss Lloyd Evans, the Principal of the Furzedown Training College, I was able to test the students in training. They made an average score of 41.

The mean variations for the various ages from nine to fourteen were successively 3·5, 5·3, 6·1, 6, 6·1, 6·3.

For the training college students it was 6.

It has been urged in criticism of the test that it measures not understanding but memory. My reply is that it measures both; and in measuring both it measures the actual intellectual gain got from the exercise. Indeed, the two factors are both inseparable and indispensable. If a reader forgets the meaning of the earlier sentences he cannot possibly grasp the meaning of the later sentences. Any word he reads may turn out to be key-word on which the significance of all that follows depends; and if the key-word, or at least its meaning, be forgotten, his reading becomes vain. Sometimes the key-word comes late and modifies the sense of all that precedes it; and if what precedes is forgotten, the point and significance of the whole piece is missed. Browning affords many instances of this kind. The key to "My Last Duchess," for example, is to be found a few lines from the end. The hint that the Duke contemplates a second marriage illumines all that he has previously

said, giving it point and purpose, and turning the poem into an intelligible whole.

Again, if the reader will refer to my absurdity test on p. 38, he will see that the absurdities all occur in the latter half, and their detection depends on an exact remembrance of what is said in the first half. Not a single sentence is absurd in itself: it is only absurd when brought into relation with the rest; and if the incompatibility is to be seen, the rest must be remembered.

Admitting, however, that it is impossible to understand the piece without also to a large extent remembering it, is it not possible to remember it without understanding it? It is possible, but so highly improbable that the possibility need not be seriously considered. It is only necessary to imagine a similar test given in Latin to an intelligent lad who knows no Latin, but can only read the words as though they were English, to see the futility of supposing a mere verbal memory to be of much service in a three minutes' test of this kind. For all practical purposes, if the subject understands the narrative, he will be able to supply the missing words; and, conversely, if he remembers the missing words, it will be due to the fact that he has understood the narrative.

CHAPTER IX

SPELLING

The usual mode of testing spelling in school is by means of an "unseen" piece of dictation. The main objection to this procedure is that the piece given is not standardised: its difficulty is unknown. And passages of prose vary so enormously in difficulty that to lay down any general rule regarding the number of mistakes per line that might reasonably be allowed at different ages is clearly impossible.

Opinions differ respecting the range of words which a child should be expected to spell. Some think that he should never be required to spell words outside his own vocabulary, and that his written composition should be the only basis on which his ability to spell should be estimated. To this view it may be objected that a child's vocabulary is a growing thing, and it would be well to anticipate a little and to teach him to spell words which he will probably want to use in the near future. Moreover, it is sometimes necessary to write from dictation, or to record in notes the sayings of others; and if the writer's own vocabulary is very limited, or restricted to certain favourite words (as in fact most people's vocabulary is), the range of his spelling vocabulary should be extended

so as to include the words that are in most common use. Thus we have two overlapping groups of words whose spelling should be known : first, those which constitute the pupil's speaking and writing vocabulary ; secondly, those most frequently used by the nation as a whole. But which are the words most frequently used ? And how many of them are we to include ?

In America these questions have been answered by Dr. L. P. Ayres, who has compiled a list of the thousand commonest words in the English language, or rather the American variety of the English language. The method employed was to examine written material of various types, such as letters, newspapers, and children's compositions, and to make a list of the words used and the number of times each word occurred. Ayres combined the results of four such studies, which comprised 368,000 words written by 2500 different persons. Some interesting facts emerged. It was found, for instance, that fifty of the words were used so frequently that they made up about half of the material examined. In order to secure a thousand words it was necessary to include words which only occurred once every 8000 words.

By testing a large number of children with this list Ayres was able to divide the words into twenty-six groups, named after the letters of the alphabet. The words in each group (the groups vary considerably in size) are supposed to be of equal difficulty. The following list comprises the words in Group N. The percentages of these words that should, according to Ayres, be spelt correctly in the grades from III to VIII are 58, 79, 88, 94,

98 and 100 respectively. Translated into English terminology it means roughly that Standard II children should be able to spell 58 per cent. of the words, Standard III 79 per cent., and so forth.

Ayres' Spelling Test

Group N.—Except, aunt, capture, wrote, else, bridge, offer, suffer, built, centre, front, rule, carry, chain, death, learn, wonder, tire, pair, check, prove, heard, inspect, itself, always, something, write, expect, need, thus, woman, young, fair, dollar, evening, plan, broke, feel, sure, least, sorry, press, God, teacher, November, subject, April, history, cause, study, himself, matter, use, thought, person, nor, January, mean, vote, court, copy, act, been, yesterday, among, question, doctor, hear, size, December, dozen, there, tax, number, October, reason, fifth.

The most useful list for English schools is Mr. Burt's, which I append—

Burt's Graded Spelling Test

Age.

5.—a it cat to and the on up if box

6.—run bad but will pin cap men got to-day this.

7.—table even fill black only coming sorry done lesson smoke.

8.—money sugar number bright ticket speak yellow doctor sometimes already.

9.—rough raise scrape manner publish touch feel answer several towel.

10.—surface pleasant saucer whistle razor vegetable improvement succeed beginning accident.

11.—decide business carriage rogue receive usually pigeon practical quantity knuckle.

12.—distinguish experience disease sympathy illegal responsible agriculture intelligent artificial peculiar.

13.—luxurious conceited leopard barbarian occasion disappoint necessary treacherous descendant precipice.

14.—virtuous memoranda glazier circuit decision mosquito promiscuous assassinate embarrassing tyrannous.

Note.—"The words assigned to any given age are those answered correctly by 40 to 60 per cent. of the age-group specified. Thus, approximately, 50 per cent. aged 10 last birthday can spell 'surface' . . . 'accident.' Therefore, to attain a mental age of $10\frac{1}{2}$ a child should spell fifty-five words, *i. e.* all to 'vegetable' inclusive (or the equivalent).

"It is not necessary that all the children should be given all the words: twenty or thirty from the appropriate consecutive ages are usually sufficient. The mental ages assigned to the earlier words are somewhat arbitrary owing to wide differences in infants' schools."

A protest should be made against the frequency with which unseen pieces of dictation are given in the higher standards of English elementary schools. Sometimes two lessons per week are devoted to it in the top class. The defence generally made is

that the children are backward in spelling; to which it may be replied that dictation is primarily a test, a means of measurement, and not a means, except indirectly, of teaching spelling. At any rate it is not the most economical means. When a child suffers from malnutrition we do not try to fatten him by weighing him twice a week.

CHAPTER X

ARITHMETIC

(*a*) The Fundamental Processes

Several attempts have been made in America to standardise tests in Arithmetic. The most widely used, the Courtis Tests, share with all the rest the defect of insularity: they constitute a grade-scale and not an age-scale, and the monetary units are dollars and cents and not pounds, shillings and pence.

Little has been done to establish norms suitable for England. In 1913 and 1914 I carried out an extensive research into the ability of London children to perform the simple fundamental processes of adding, subtracting, multiplying and dividing. The results of this research were published in the *Journal of Experimental Pedagogy* for December 1914 and March 1915. The tests were repeated and the standards broadly confirmed by Mr. P. L. Gray, H.M.I., in the schools of Leeds. The copy of the *Journal* (June 1915) which contained Mr. Gray's account of his experiment also contained a criticism of my tests by the editor, Professor J. A. Green. At the meeting of the British Association at Newcastle in 1916 the Report of the Committee on the Mental and Physical Factors involved in Education dealt with Norms in Mechanical Arithmetic and gave an account of investigations in

London and Sheffield on similar lines to my own. The tests used (they were devised by Mr. Cyril Burt) were better than mine. Mine suffered from the defect of being too brief. Five minutes were allowed for working twenty-four sums, and in nearly every school tested some of the older children finished the paper before the time was up. If the paper had been longer their score would have been higher.

In the autumn of 1919 I revised my original tests and administered them to a large number of London children. They differ in important respects from the 1914 tests. They are both simpler and more searching. The time allowed is three minutes instead of five, and none of the children tested were able to complete the papers within the time. A yearly age-scale is substituted for a half-yearly age-scale. Finally, the numbers have been so selected as to secure the greatest variety compatible with leaving all the sums of approximately equal difficulty. If a multiplication table be constructed up to 9 times 9, and a similar table for each of the other processes, every item in those tables will be found represented somewhere in the examples set; generally twice, once in the first half and once in the second.

Addition: Three Minutes

64	35	82	78	46	27	3
16	29	63	92	58	56	98
31	40	9	50	71	93	51
98	78	14	63	42	16	43
2	51	23	17	39	72	19
75	47	65	89	4	80	65

70	44	86	39	81	24	16
28	8	97	85	50	69	92
76	57	42	16	72	92	35
52	26	50	46	24	5	48
13	41	13	9	68	74	70
89	93	72	53	73	38	13

29	36	7	51	89	40	21
81	90	24	85	73	49	5
6	54	19	24	15	58	18
47	18	57	39	24	24	58
58	63	26	66	16	79	73
72	59	38	70	3	63	64

92	16	27	34	53	69	75
15	37	31	16	27	92	43
70	4	96	52	69	28	82
48	48	80	79	15	15	68
29	95	58	48	34	73	7
36	24	27	8	78	40	94

Subtraction: Three Minutes

69152	80031	68703	54218
48729	63175	37956	49221
17690	91435	62098	76112
7948	23256	34089	57346

85694 2787	60321 14359	70592 26819	80941 53437
78580 29395	60211 43716	81427 42078	93505 16693
74816 53617	92240 59830	80134 46965	75702 28139
78265 2495	70124 65139	80367 20548	94521 16973
35172 19438	59203 47689	69375 7089	49315 19782

Multiplication: Three Minutes

273905 4	360197 7	591472 5	465239 6
629175 2	196704 9	297836 3	519824 8
148637 4	254879 7	368051 5	780147 6

308465 2	328547 9	541783 3	736582 8
612549 7	598162 4	561923 6	726493 5
598263 9	793428 2	439716 8	625097 3
180493 8	641857 5	834729 4	582736 7

Division : Three Minutes

4/26930	7/66759	5/48175	6/44957
2/36514	9/53412	3/28103	8/58849
4/57792	7/13026	5/82947	6/33802
2/12978	9/16743	3/24861	8/85390
7/59304	4/21159	6/52298	5/63772

9/24419	2/18758	8/39857	3/55378
5/19238	7/89292	9/34263	4/78295

The examples were graphed or cyclostyled and distributed face downwards, one process at a time. Three minutes were allowed for working, and one mark allowed for each sum that was absolutely correct.

The following norms were arrived at—

Number of Sums right in Three Minutes

Age.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.
Addition . . .	3	4	5	6	7	8
Subtraction . .	2	$3\frac{1}{2}$	$4\frac{1}{2}$	$5\frac{1}{2}$	$6\frac{1}{2}$	7
Multiplication . .	$1\frac{1}{2}$	3	$4\frac{1}{2}$	$5\frac{3}{4}$	$6\frac{3}{4}$	$7\frac{3}{4}$
Division . . .	1	$2\frac{3}{4}$	4	$5\frac{1}{4}$	$6\frac{1}{4}$	7

If we mark the papers in another way and instead of counting the number of sums right count the number of operations right, we shall get a more exact score, for examples partly correct would score marks. By operations I mean processes of the kind tested. For instance, in the first addition example there are ten addition operations, in the first subtraction example five subtraction operations. For multiplication and division the corresponding numbers are six and four. The advantage of giving the norms in operations per minute, as in the following table, is that in applying a rough test any examples may be set by the teacher, provided he makes a little allowance for the size of

the sums, and the time taken in writing the figures and in passing from one sum to another.

Number of Operations per Minute

Age.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs,	14 yrs.
Addition . . .	12	16	20	24	27	30
Subtraction . .	4	6	8	10	12	13
Multiplication. .	4	7	10	12	14	16
Division . . .	2	4	6	8	9	10

When reduced thus to operations per minute the results of my 1919 investigation confirm the results I obtained in 1914.

Comments on the Addition Test.—The improvement with age is only partly due to a more facile application of the same method : it is mainly due to a change of method, to a gradual supersession of habits of a lower order by habits of a higher order. The lowest habit of all is that of adding by units ; the highest, that of adding *en bloc*. For example, in adding the series $8 + 5 + 9 + 2 + 7 + 1$ the pupil working by the former method would start $8 + 1 + 1 + 1 + 1 + 1$, etc., while the pupil working by the latter would say $8 + 5 = 13$ as a mere matter of rote memory. Between these two extremes there are various intermediary stages. Some, for instance, decompose the addends in order to form tens. Thus $8 + 5 = 8 + 2 + 3 = 10 + 3 = 13$. Adding the next item, $13 + 9 = 13 + 7 + 2 = 20 + 2 = 22$, etc. Some other devices used to avoid counting by units are also based upon the decomposition of the addends into groups that can be added without further reduction. Then there is the method of arriving at the result by noting how far it falls

short of a readily ascertained sum. For instance, $13+9=13+10-1$; and $8+7=8+8-1$. Lastly, there is the method advocated in some schools of searching along the column for numbers which form tens. Thus in the series given above the 8 and the 2 would be coupled, and also the 9 and the 1. This is a useful device when the paired numbers come together; but to link them when widely separated is to increase unnecessarily the mental strain, and to run the risk of omitting or duplicating some of the addends.

There can be little doubt therefore that the most efficient method is that of adding the items *seriatim* and *memoriter*—as they come and by the addition table. Advantage should, of course, be taken of any obvious ten-group that would appear, but there should be no hunting for it. It has been maintained by some that to add $4 + 9$ as $4 + 10 - 1$ is more "intelligent" than to say straight off $4 + 9 = 13$. But why should it be so? There is an assumption that $4 + 10$ is known, while $4 + 9$ is not. Logically the proposition $4 + 9 = 13$ rests on the number system, 1, 2, 3, 4, etc., and counting by units is, in a sense, the one and only "intelligent" method—the one and only method that lays bare the ultimate ground of the proposition. But it is assumed that the rationale of adding is clearly understood by the pupils; and the question under discussion is not the most explicitly logical but the most expeditious mode of adding. As a step towards memorising $4 + 9 = 13$, the operation $4 + 10 - 1$ may be permitted; but the process is not complete from the point of view of practical efficiency till it is short-circuited into $4 + 9 = 13$.

The number of items to be thus memorised is not great. For if $4 + 9 = 13$ be learnt, the sums of all such combinations as 14 and 9, 29 and 4, etc., are easily inferred therefrom.

The time taken to work the addition paper gives a clue to the method used. I have applied the test to a few children without imposing a time limit. The children handed in their papers as soon as they were finished, and the time taken was recorded thereon. Later on each child was privately asked to work through one of the sums orally, in order that the method adopted might be ascertained. The results confirmed the view that both speed and accuracy depend upon the grade of habit upon which the process of adding is based.

The danger of forming bad habits of computation is far greater in addition than in multiplication. In working the latter there is little temptation to count. The veriest tyro cannot fail to perceive the enormous labour-saving advantage of remembering that $6 \times 8 = 48$ over the counting of six eights; but the advantage of memorising $6 + 8 = 14$ is not so obvious when the result can readily be arrived at by a mechanical drumming with the fingers. It is indeed a notorious fact that a certain percentage of educated adults count on the fingers —when nobody is watching them.

Comments on the Subtraction Test.—The most salient feature of the subtraction results is the evidence afforded of the superiority from a practical point of view of the method of equal addition over the method of decomposition. In my account of the 1914 investigation I gave abundant evidence to this effect, evidence which is further strengthened

by my recent investigation. When the subtraction score was in any school markedly below the scores for the other three processes I guessed that the method of decomposition was taught at that school; and my guess was always right. Let me take as an example one of the best schools tested. Situated in a very good residential district, it is attended by children of exceptional advantages. The girls' school is noted for the general thoroughness of the work. Now observe the difference in the results obtained in the two departments both at the lowest age and the highest age tested.

		+	−	×	÷
9 yrs.	Boys	3·5	4·1	3·1	2·8
	Girls	7·3	3·8	6·3	4·7
14 yrs.	Boys	6·8	7·6	7·8	7·2
	Girls	10·2	8·7	11·4	10·3

An examination of these figures will show that the girls' school, where the method of decomposition is adopted, is, in spite of the excellent teaching labouring under a heavy disadvantage as compared with the boys' school, where the method of equal additions is taught. The figures also illustrate the fact that although the advantage of one system over the other tends to diminish with the age of the children, the advantage is never lost—not at any rate during the elementary school period.

But the disparity in the practical efficacy of the two methods is really greater than would appear from the above account; for the examples set having few or no noughts in the minuends failed to bring out the more glaring disadvantages of the decomposition method. When, however, such an example

as 40,000 — 197 is set, the ineptitude of the method is more strikingly revealed. Even in the highest classes of the decomposition schools the time taken to work such an example is excessive, and the degree of accuracy abnormally low. And yet it is examples of this type (where the minuend is a round number) that are of most frequent occurrence in everyday life. This is obviously so in the case of money. We want change from a five-pound note, a sovereign, a half-sovereign, or a shilling, and not from such a collection of coins as is represented in £4 13*s.* 7¾*d.*

The reason for the inferiority of the decomposition method is not far to seek. In the equal addition method the compensation is made—accounts are squared—at the very first number dealt with after the minuend has been disturbed. In subtracting 37 from 85, after taking 7 from 15 the disturbed relationship of difference between minuend and subtrahend is immediately restored by increasing the 3 tens to 4 tens. In the method of decomposition, however, it is the 8, the second figure dealt with, that has to be changed to restore the balance. If the minuend figure is zero the balancing of accounts is still longer deferred. In a phrase, the main secret of the difference lies in the dispatch with which accounts are settled. One is "cash payment," the other "credit." And the postponement of the compensating act increases the chances of its fulfilment being forgotten. But this is not the only point of difference between the two methods, for there is a further disparity in the area of disturbance. When the figure in the minuend represents a smaller number than the corresponding figure in the subtrahend it is necessary to disturb

the minuend (method of decomposition) or both minuend and subtrahend (method of equal addition). In the latter case it is never necessary to disturb more than two figures; in the former it is always necessary to disturb two, and sometimes many more; and for a young child to bear the many changes in mind is no easy task.

The disadvantage of the decomposition method is not of course limited to pure subtraction sums: it vitiates all exercises into which subtraction enters. Long division, for instance, is, as I have abundantly tested, performed in decomposition schools with difficulty and with dubious accuracy.

And yet the decomposition method is apparently taught in about two-thirds of the London schools. What is the reason for its popularity? It is not the method we learnt in our youth; it does not seem to be the method adopted by the adult, even when he has learnt that method at school. The reason is to be found in its greater intelligibility. It is easier for a child to understand the decomposition of numbers than to grasp and apply the principle that the difference between two numbers remains unchanged if the same number be added to both. It is therefore the favourite method in the infant school, and the senior school follows the lead. But granted its greater intelligibility, its practical efficiency is not encouraging. The younger the child the more is he shackled by the inferior method. Unless, indeed, we assume that some of the older children in the decomposition schools discover in some indirect way the equal addition method and use it in preference. It is not often that one finds a class of older children all of whom

practise the decomposition method. Do we not in any case pay too great a price for a doubtful boon? If at first the child sees the rationale of the process of decomposing the minuend, he soon gets to perform it automatically. The "intelligence" supposed to be concerned is a temporary illumination only. Indeed for pure practical efficacy it is better that the rationale should not, at the actual time of working, be thought of at all. The pupil should confine his efforts to a rigid application of the rule. I have frequently observed that when teachers in training are asked to work a subtraction sum and explain the steps, they frequently give the right reason but the wrong answer.

It may be pointed out that an intelligent application of an arithmetic rule does not necessarily mean a scientific knowledge of the underlying principles. A child may learn to walk and to put the power to intelligent uses without knowing anything about the mechanism by which walking is achieved. So may he learn to work subtraction by rule of thumb, and be able to apply it quite intelligently to the practical purposes of life. He may later on study the physiology of walking, or the logical basis of the rule; but there is no more reason to think that he will compute better in the latter case than there is for thinking that he will walk better in the former. This is not a plea for the mechanical teaching of subtraction; but it is a plea for regarding subtraction as primarily an instrument to be placed in the hands of the young pupil for the purpose of solving certain problems of actual life. If he can understand the reasons for the steps taken, well and good. If not, he should for the present use it

without understanding it. Indeed I have long contended that during the last year or so of an elementary pupil's schooling he should be taught the underlying principles of all the rules he has learnt. Attempts should, of course, be made to render the rules intelligible at the time of learning; but the teacher should be in the main content if the pupil can use these rules in concrete cases. A critical examination of "long division," for instance, is a valuable exercise for a lad of 13—far more valuable than the senseless manipulation of symbols which often passes for algebra. The mathematics for the last year of the school life of an elementary pupil should include the assimilation of all the undigested material in the whole arithmetic course, if that course is, as it should be, regarded as a systematic study of the principles of numbers.

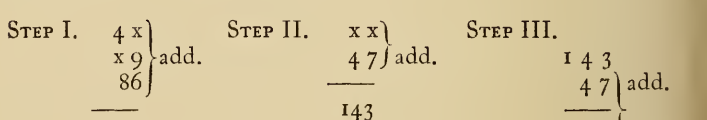
It has frequently been asserted that there is a third method of teaching subtraction—the method of complementary addition. But this is not, like the two methods just dealt with, a device for meeting the difficulty of "borrowing": it is an alternative way of looking at the process of subtraction itself. $16 - 7$ may mean either: (*a*) what is left when 7 is taken away from 16? or (*b*) what must be added to 7 to make 16? If the latter view be adopted then subtraction is regarded as complementary addition—as a solution of the equation $a + x = b$.

It may be remarked about this form of subtraction that it is not taught as a general process in any of our schools, at least not to my knowledge. It is true that complementary addition sometimes appears in a hybrid form. Thus $43 - 26$ is some-

times worked in this way: 6 from 10 leaves 4, 4 and 3 are 7, etc. But such roundabout methods, in which two steps are taken where only one is necessary, are not to be commended.

Complementary addition pure and simple, combined with equal addition as a "borrowing" device, is advocated at some of the Universities, especially where much work in logarithms has to be done. Instances are known where greater speed and accuracy have resulted from a change from the subtractive to the additive attitude.

The additive view is strongly urged by the few head teachers who have tried it. They think that the subtraction method should be discovered by the child: the steps being indicated by the following examples—



The "carrying" or compensating step is naturally adopted, with no need for more explanation than the "carrying" in simple addition.

It is not for me to judge this method on *a priori* grounds, but its advantages are sufficiently obvious for us to declare a true bill in its favour and give it at least an experimental chance.

On the Importance of Tables.—It has already been shown that the facility in working addition and subtraction mainly depends upon a ready knowledge of the addition table.

The complete table may be written in this form—

1+1= 2
2+1= 3 2+2= 4
3+1= 4 3+2= 5 3+3= 6
4+1= 5 4+2= 6 4+3= 7 4+4= 8
5+1= 6 5+2= 7 5+3= 8 5+4= 9 5+5=10
6+1= 7 6+2= 8 6+4= 9 6+4=10 6+5=11 6+6=12
7+1= 8 7+2= 9 7+3=10 7+4=11 7+5=12 7+6=13 7+7=14
8+1= 9 8+2=10 8+3=11 8+4=12 8+5=13 8+6=14 8+7=15 8+8=16
9+1=10 9+2=11 9+3=12 9+4=13 9+5=14 9+6=15 9+7=16 9+8=17 9+9=18

It will thus be seen that there are 45 results to be memorised—45 habits to be fixed—it being understood that each of the above items represents four processes. Thus $8 + 4 = 12$ should also be memorised as $4 + 8 = 12$, $12 - 8 = 4$ and $12 - 4 = 8$; not of course as independent processes, but as necessarily implied in the one formula $8 + 4 = 12$.

The same remarks apply to the multiplication table, except that there are only 36 items to be learnt.

There are some educationists who contend that the tables should not be memorised. In so saying they do not mean that they should not ultimately be memorised; but rather that no conscious effort should be made to memorise them. The results should be arrived at either—

(*a*) By building them up afresh each time, or
(*b*) By referring to a table book.

If they are continually being built up afresh, any intellectual value such a process may originally possess soon disappears. It sinks to the level of the merest mechanical work. And if this method is to be applied to the multiplication table, logically it should be applied to the addition table as well, and counting by units should always be encouraged.

If, on the other hand, it is merely meant that the

results should be calculated *ab initio* each time until they are fixed in the memory, experience shows that the mere habit of doing so tends to form a stronger tendency to start the process than to recall the result. Older children, for instance, who count on their fingers delay indefinitely the counting by groups. No effort is made to memorise the results, and in consequence they elude the memory. Indeed a far shorter way of reaching the goal is afforded by the second alternative, that of using a table-book—assuming, of course, that the construction of the tables is understood. Whenever, for instance, the product of 7 and 8 is needed the table is looked at. Here the mind attends exclusively to the result, 56, and is not absorbed in attending to the process. Mr. Winch has experimental evidence to show that if a large number of examples involving the use of a specific table are worked rapidly by the pupils who have this table in front of them, it is actually memorised better than if a conscious effort is made to memorise it without working examples.[1]

It seems clear that progress in mathematics is made possible by assuming the results of previous processes and using these results as stepping-stones to still higher results.

On the Best Method of Memorising the Tables.—The addition table has generally been left to look after itself, but the multiplication table has always received a certain amount of attention. In bygone days it was systematically memorised by frequent simultaneous repetition; and even at present the chanting of the tables, although less widely adopted,

[1] See also Kirkpatrick: "Memorising versus Incidental Learning," *J. of Educ. Psychol.*, v. 7, 405–412.

is almost the only means that is employed. But this chanting of the tables is open to several objections.

Memorising of all kinds depends upon the fixation of a habit-series; and the limits of the series should be clearly defined. Each formula, such as $4 \times 7 = 28$, constitutes a self-contained system, and it should be so memorised as to be completely usable without reference to preceding formulæ. In other words no unnecessary associations should be set up. To associate by rote memory 4×5 with 4×6, and 4×6 with 4×7, etc., is a superfluous, if not an injurious, bit of mental mechanisation. Chanting tends to establish these useless associations.

Another objection is that the speed of this simultaneous repetition is far too slow for the economic fixation of habit. The effect of speed upon mechanisation, although not generally recognised, is considerable. If, for instance, a passage of poetry has to be memorised so as to render its repetition automatic, the repetition of the lines at maximum speed has been found, in my own case at least, to diminish the number of repetitions necessary. It has probably something to do with the span of attention.

A third objection to the simultaneous chanting of tables is based upon the liability of the attention to wander during the repetition. Attentive repetition is far more efficacious than the inattentive kind.

Finally, there is the objection that may be urged against all kinds of simultaneous class work; that is, that it makes no allowance for individual differences in the mode and rate of learning.

There are experimental grounds for believing

that the method of individual muttering—a method which unfortunately seems to "get on the nerves" of some teachers—is considerably more efficacious than the method of concerted repetition.

There are two methods of memorising the tables which I have reason to think would prove effective—methods which are not exclusive but supplementary. *Both* might be tried.

(1) Take, say, two items per diem in the addition table, such as $7 + 8 = 15$ and $4 + 6 = 10$; and two per diem in the multiplication table, such as $7 \times 8 = 56$ and $4 \times 6 = 24$. If this be systematically done, and past work frequently revised, the whole will be learnt in less than five weeks. If only one of each be taken per day the whole can be mastered in less than three months.

(2) Learn by applying. Put, for example, the 7 times table before the boys and let them work *very rapidly* a large number of sums involving multiplication by 7. Of the two methods this is probably the better.

On the Importance of Practice, and the Claim of the Individual.—That the principles of number should be intelligently taught; that they should be recognised by the pupil as rooted in the experiences of everyday life; that they should be learnt and applied with understanding; that they should frequently be presented in novel combinations—these are matters upon which there is now no divergence of opinion. But when we ask the questions: Should every exercise be given in concrete form? Should the problem dominate the arithmetic lesson? we get a variety of answers.

There are many who believe—there are more

who wish to believe—that sufficient practice in the mechanism of computation is gained by working problems and problems only. But whatever opinion one may have held ten years ago on this matter, if one has followed the recent development of mathematics in the elementary schools one cannot help being forced to the conclusion that such practice is insufficient. Ciphering, in its rudimentary forms, is so useful an art that proficiency therein is justly regarded as one of the essential aims of an elementary school, and to discover the most economical means of achieving that aim, without doing violence to the pupils' instincts and interests, will ever be one of the central and vital problems of teaching.

It is not fair to argue that since the excessive and exclusive grind at mechanical arithmetic which characterised the period of payment by results was distasteful to the pupils, mechanical arithmetic is in itself distasteful. Indeed, many children share the opinion of the little girl of my acquaintance who says that she likes sums but does not like sums about John.

A noticeable feature of the present arithmetic course is the absence—or at least the infrequence—of the pure practice or drill lesson. From having all the lessons practice lessons we have come down to none. The custom of wrapping up numbers in abundant verbiage has become so inveterate that the mere sight of a naked number gives some of us a sort of shock. Some time ago I was reproached by a teacher for asking a little girl in an infant school to add 2 and 3. All departure from the concrete has come to be regarded as wicked. We have in consequence an ounce of arithmetic to a

pound of padding. I have seen a teacher spend ten minutes over this little question in mental arithmetic: "If 80 birds sat on a tree, and 30 of them flew away, how many would be left sitting on the tree?" Laboriously she wrote it on the board, and persistently she checked all attempts at answering until she had explained the situation to the point of boredom.

Another factor unfavourable to progress is the non-recognition of the essential heterogeneity of any collection of children, however carefully chosen. Any seeming homogeneity in a class is both superficial and temporary. However much the units may resemble one another in their present attainments they will differ enormously in their capacity for work, and consequently in their rate of improvement. If they appear like one another to-day, they will appear unlike one another to-morrow. Although this fact has been clearly demonstrated by others,[1] I have taken steps to verify it for myself. Three classes in a girls' school were allowed to work through the exercises in their arithmetic text-books at their own pace for half-an-hour per day. The results were as follows—

	Class A	Class B	Class C
Number in Class	50	46	42
Approximate age of Children	10	12	13
Number of half-hour Lessons	27	26	23
Average total number of sums worked correctly	206	145	181
Highest Score	390	250	514
Lowest Score	95	70	24
Mean Variation	57	34	87

[1] See, for instance, Search's *Ideal School*, pp. 29 and 33.

It would be difficult to find a school more carefully organised—a school where the children in a class were nearer the same level—and yet the variability in their rates of working is seen to be enormous. In the highest class the best girl was able to work more than 21 times as fast as the worst; and although this amount of disparity is exceptional, rarely will it be found that the fastest in the class does not work at least three times as rapidly as the slowest.

It will be seen that, taking the three classes together, on an average 7 sums per child were worked correctly in half-an-hour. Five girls in the top class were able to do more than double this average. It must not be thought that the exercises worked were of an easy type, or were of a merely mechanical nature; they were continuous examples from McDougall's *Suggestive Arithmetics*; and these books are above, rather than below, the average in difficulty.

A careful investigation of the methods of teaching arithmetic at present in vogue in elementary schools has convinced me that the most serious and prevalent defect is the excessive use of the blackboard, both for setting exercises to be worked by the class, and for exposition—for explaining, and exemplifying, and correcting.

When we consider that it is rare for a class working from examples written on a blackboard to get through more than four sums during a lesson of forty-five minutes (that is, half as long again as the lessons referred to above) it becomes obvious that the individual scholar is not working at anything like his normal pace. I do not go so far as to urge that a child should *always* work at the highest

pressure, but I do submit that he should *sometimes* do so.

It is often the practice of the teacher to write an example on the board, and set the whole class to work it on paper. After all have finished (note the waste of time on the part of the brighter children) the sum is marked. If about one-third of the class gets it wrong, the teacher, as a rule, works it himself on the blackboard, or gets the class to work (or to seem to work) it with him. This is idle time for two-thirds of the class; and it is not the best method of correction for the other third. More often than not the mistakes are due to carelessness on the part of the pupil, or to an imperfect knowledge of the tables; and effective correction depends on individual effort rather than on blackboard explanation.

An arithmetic lesson is occasionally taken up by the teacher working one or two examples on the blackboard with the class, the children afterwards copying them out in their exercise books. The concerted appearance of this work is illusory. The work is really done by the teacher and a few of the more alert pupils. The copying out is of little value except as a relaxation from the boredom of watching other people work.

These practices are by no means universal, nor do they imply that the instruction is mechanical and perfunctory. In making the criticisms it is but fair to record my conviction that earnest efforts are almost universally being made to vivify the instruction and to bring it into line with modern educational ideals, and that whatever faults exist are due not to lowness of aim, but rather to a misconception

of how a high ideal may best be approached. The spirit of the *Suggestions to Teachers* and of the Report of the L.C.C. Conference on Arithmetic has permeated the majority of schools, and is doing incalculable good; but at the same time many sound practices of the past have sometimes been forgotten, and the new conditions that are developing with the gradually diminishing class and the more rapid movement of pupils have not been met by a corresponding change of method. That change may briefly be described as a progress from class teaching to sectional teaching, and from sectional teaching to individual teaching—in fact from larger to smaller units. The ideal teaching is, of course, individual teaching, and Rousseau, in his *Émile*, assumes as a principle: one teacher, one pupil.

It will be seen that "blackboarditis" (if I may be permitted to call it so) arises from too ardent and sanguine a desire to preserve the unity of the class—from the belief that the individuals forming this unity should work at the same rate, and progress at the same rate. One teacher with whom this matter was discussed not only held this doctrine, but strictly maintained that the pace should be that of the slowest pupil. This indeed seems to be the only logical form of the doctrine. It is clear that the average pupil (if there be such) cannot set the pace, as that would leave about half the class in the lurch. In actual practice a small number is recognised as forming no real part of the body of the class, and is labelled the "tail end." The "tail end" is left out of account, and the pace is virtually fixed by the slowest pupil among the remainder.

But it is clear from what has already been said that the doctrine of the homogeneous class ought to be abandoned, and the teachers should devise means of securing the maximum of healthy effect from each individual child.

That it is necessary to devise such means is obvious when we consider the size of the class and the limited time and energy at the disposal of the teacher. He cannot possibly devote the necessary amount of personal attention to each child. The solution of the difficulty lies in the delegation of his powers, and in fostering in his pupils a sense of responsibility. We have not yet discovered the extent to which we can trust the pupils. By adopting a general policy of mistrust, by never allowing a child to mark his own, or even another child's, exercises, by making no child responsible for anybody's conduct or progress but his own, by retaining all corrective and coercive powers in the teacher's hands, we gain certain advantages: we simplify matters, we minimise the likelihood of abuse of authority, and we cultivate in the pupils the virtue of obedience. But we lose much more than we gain. We fail to secure normal and healthy progress, we fail to foster respect for an internal as distinct from an external authority, we fail to cultivate the power to rule wisely as a balance to the correlative virtue of obeying wisely, and we sacrifice the brilliant and the stupid to the mediocre. The surprise with which we view the success of the prefect system among elementary school children is itself a mark of our traditional mistrust. The possibilities of self-government and self-culture among school children will, I believe, when fully

realised, provide the key to the solution of a large number of the difficulties which press upon the teacher at the present time, and which seem to put an abnormal strain upon his nervous system. I am not here concerned with showing how precisely this delegation of work and responsibility may be effected in the arithmetic lesson (each teacher can best discover this for himself), but with suggesting the direction in which the ideal of individual development probably lies.

Recommendations.—In view of the facts and arguments set forth above, I venture to make the following definite recommendations—

(1) That the tables, both addition and multiplication, be by some means or other fixed in the memory early in the arithmetic course.

(2) That the simultaneous repetition of the tables be superseded by individual learning, or better still, by their application to examples to be worked rapidly.

(3) That seriatim repetition be discarded after the structure of the tables is understood.

(4) That adding by tables be the final objective in practising addition, and that adding by units, or by partial groups, or through any roundabout device, be regarded as a habit of a lower order, to be abandoned as soon as habits of a higher order can be engendered.

(5) That speed of adding be insisted on as a means of pressing forward towards the higher habits.

(6) That the method of equal addition be universally taught as the practical method of working subtraction.

(7) That the method of decomposition be regarded, if taught at all, as a means of showing the correctness of the result arrived at by the usual method.

(8) That at least one pure practice lesson be given per week.

(9) That speed as well as accuracy be aimed at in the practice lesson.

(10) That the terminal examination in arithmetic contain at least one straightforward abstract sum.

(11) That each class be frequently practised in the work of all the lower classes.

(12) That means be adopted to secure the progress of each pupil at his own natural rate.

(13) That the blackboard be not used for setting out examples when text-books are available for that purpose; nor for working sums which could easily be worked by the majority of the class; nor for correcting errors due to mere carelessness. (The blackboard has, of course, its legitimate use for class and sectional teaching; it is only when it becomes a means of preventing individual effort that its use is open to objection.)

(14) That the practice of copying in the exercise books examples worked out on the board be discarded.

(15) That much of the responsibility of marking exercises be, with due reservations and precautions, delegated to the pupils.

(*b*) Simple Oral Arithmetic

Many a teacher has felt the need of some simple means of estimating, with a reasonable degree of

precision, the arithmetical attainments of young children. The tests that have hitherto been published are intended for older children, for children who are able to work sums on paper. They cannot measure the beginnings, and are consequently of no use in dealing with children in infant schools.

To meet this need I have used the following tests in addition and subtraction—

One Minute Oral Addition Test

(1) 1 + 2	(11) 4 + 4	(21) 7 + 5
(2) 4 + 1	(12) 5 + 2	(22) 8 + 3
(3) 2 + 2	(13) 6 + 4	(23) 4 + 9
(4) 2 + 4	(14) 1 + 8	(24) 6 + 8
(5) 3 + 2	(15) 3 + 7	(25) 7 + 6
(6) 4 + 3	(16) 6 + 3	(26) 9 + 8
(7) 2 + 5	(17) 2 + 6	(27) 9 + 6
(8) 5 + 4	(18) 5 + 5	(28) 8 + 7
(9) 3 + 5	(19) 7 + 2	(29) 5 + 9
(10) 8 + 2	(20) 4 + 6	(30) 7 + 9

One Minute Oral Subtraction Test

(1) 2 — 1	(11) 8 — 2	(21) 11 — 2
(2) 3 — 2	(12) 7 — 5	(22) 10 — 6
(3) 5 — 1	(13) 8 — 3	(23) 12 — 3
(4) 6 — 2	(14) 7 — 4	(24) 11 — 6
(5) 5 — 3	(15) 9 — 3	(25) 12 — 5
(6) 2 — 2	(16) 8 — 5	(26) 13 — 4
(7) 7 — 2	(17) 10 — 4	(27) 15 — 9
(8) 6 — 4	(18) 9 — 5	(28) 14 — 6
(9) 7 — 3	(19) 10 — 3	(29) 17 — 8
(10) 6 — 3	(20) 9 — 4	(30) 16 — 7

Each child is tested individually and in isolation. He is asked the question, "One and two?" and as soon as he answers it he is asked the next, "Four and one?" and so on. He is not allowed to proceed until he has given the right answer. The number of questions correctly answered in one minute gives the score.

The same method is adopted for administering the subtraction test.

A little preliminary questioning is desirable to put a child at his ease, and to see that he knows exactly what is required of him.

The following norms are based on the results obtained by myself and others in the application of the tests to about ten thousand boys and girls within the last six years.

Age.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.
Addition Score . . .	6	10	14	18	22
Subtraction Score . .	4	7½	11	14½	18

The mean variation seems to be nearly the same for each age group and for each process, and amounts to between three and four.

The girls are able to add a trifle better than the boys, but in subtraction there is no appreciable difference between them.

With the very youngest children the score is low, not merely because they are slow in computing, but because their knowledge is limited to the first few numbers in the scale. They cannot answer the whole of the questions even if they are given unlimited time. When, however, older children score few marks it is because they add by units. This points to the expediency of memorising the

addition table early. The answering in a few schools, however, convince me that there is a danger of premature memorising. $2 + 5 = 10$, for instance, is a type of answer by no means infrequent. This confusion between the addition and the multiplication tables is scarcely possible if the child is familiar with the make-up of the first ten natural numbers.

In subtraction the question that seems to present the greatest difficulty is the sixth ($2 - 2$).

As the oral examination of a large number of pupils took a considerable time it was hoped that a saving might be effected by setting the same test in written form. It was found however on actual trial that even with the older children the written test was a less delicate measure of arithmetical ability. The questions were typed and set to a number of children between nine and ten years of age, in the form given above with a blank space for the answer. The children between nine and a half and ten did 20 per cent. worse than at the oral examination, and the children between nine and nine and a half did 30 per cent. worse. The difference would, it is reasonable to assume, be increasingly great as we went down the scale of age. It was therefore considered wiser, in dealing with such rudimentary work, to adhere to the oral form throughout.

I have endeavoured to find the difference between the achievements in good and poor schools, the terms good and poor being used in a social sense only. Instead of comparing two schools, the best and the worst—a plan accompanied by obvious risk—a group of very good schools was compared with a

group of very poor schools. The difference proved to be greater than I had anticipated. In both addition and subtraction the good schools were at least one year in advance of the age-norms, and the poor schools at least one year behind. There was, in fact, a difference of more than two years between the achievements of the two social classes.

I also found distinct evidence in support of what seems to be a general rule: the higher the social status the more favourably do the girls compare with the boys. In good neighbourhoods the girls compute a little better than the boys, in poor neighbourhoods a little worse.

(*c*) Arithmetical Devices

This test aims at discovering how many of the "rules" commonly taught in the elementary schools have been mastered. If the pupil shows that he knows the rule and has merely made a slip in computation the sum is counted right. Each correct sum scores one mark.

Arithmetical Devices (*unlimited time*).

(1) 658×204.
(2) $95567 \div 53$.
(3) £2 14*s*. $5\frac{1}{2} \times 26$.
(4) £23 11*s*. $9\frac{1}{2}d. \div 17$.
(5) Calculate the cost of 15 tons 3 cwt. 2 qrs. at £1 5*s*. 6*d*. per ton.
(6) $\frac{3}{4} + \frac{5}{6}$.
(7) $2\frac{2}{5} - \frac{3}{4}$.
(8) $\frac{3}{8} \times \frac{4}{5}$.
(9) $\frac{5}{8} \div \frac{3}{4}$.

(10) ·1 × ·1.
(11) ·6 ÷ ·003.
(12) 5 : 12 :: 9 : x.
(13) Find the average of 5, 11, 7, 2, 0, 19.
(14) What is the value of £0·168?
(15) Find the simple interest on £650 for 2 years at 5 %.

It will be observed that where a simple example will suffice to indicate whether the pupil knows the rule the simple example is adopted. No. 5 and No. 12 proved the most difficult; due no doubt to the fact that within recent years "practice" has fallen into discredit, and proportion has given place to the method of unity.

Age.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.
Score . . .	½	3	5	6½	8	9½

(*d*) Applied Arithmetic (*Problems*)

To develop the capacity to apply the principles of numbers is after all the final aim of the teaching of arithmetic; and the extent to which the pupil can bring his knowledge to bear upon the common affairs of life is the real measure of his understanding. Hence the importance which we now-a-days attach, and rightly attach, to "problems." It is regarded as questionable wisdom to reduce all problems to types and to teach them under the old system of rule and example. The root objection to this plan is that the pupil, recognising a particular problem (often by illegitimate signs) as belonging to a certain group, is liable to work it by rule of thumb rather than the direct application of basal

principles. It prevents him, in fact, from reasoning the thing out. If, however, pupils are to be trained at all in the solution of problems, it is necessary to call their attention to the underlying principles; directly or indirectly they must be taught to recognise a certain sameness in the examples, and to do this is to systematise and classify. Take for example the 19th problem in the test given below. This is the same as asking: "What are the two numbers whose sum is 5 and whose difference is 2?" Or again, "A bottle and a cork cost $2\frac{1}{2}d$; the bottle cost $2d$. more than the cork; what did the cork cost?" This, too, is cousin german to No. 10. The real danger lies not in the recognition of the type, but in working successively so many examples of the same type that the work becomes mechanical—a danger not difficult to avoid.

At any rate, whatever objections there may be to the teacher's classifying problems for his pupils, there is a distinct advantage in his classifying them for himself. He should be familiar with the common applications of arithmetic, and should be careful both in teaching and testing to vary his types. He can never perhaps exhaust the types; but he can readily make a catalogue of the commonest of them. The test below is, in fact, just such a catalogue. It comprises twenty fundamental types of problems reduced to their simplest forms. It will be observed that the actual calculations are extremely easy; for it is not computation that is being tested, but merely the capacity to apply principles. Hence a slip in counting is overlooked when the papers are marked, and every problem solved by the right method is allowed to score one.

The problems are arranged in order of increasing difficulty—an order which differs from my preconceived notion of their difficulty, and which was discovered by actually applying them to a large number of children.

APPLIED ARITHMETIC TEST (*unlimited time*)

1. If there are 100 apples on a tree and the wind blows down 17, how many are left on the tree?

2. Two hundred oranges are put into 5 baskets so that each basket has the same number of oranges. How many are there in each basket?

3. If a man earns £5 3*s.* 6*d.* in a week, how much will he have earned at the end of a month?

4. I want to buy a book which costs 7*s.* 6*d.*, but I have only 1*s.* 8½*d.* in my pocket. How much more money do I need?

5. John pays 8*s.* 4*d.* for 5 lbs. of butter. What would Henry have to pay for 7 lbs. of the same butter?

6. If eggs sell at 3 for 2*d.*, what will 2 dozen cost?

7. One day two boys earn ten shillings between them by carrying trunks from an hotel to a railway station. One boy carries 5 trunks during the day and the other 7. How ought they to share the money?

8. What is the least number that must be added to 1483 to make it exactly divisible by 16?

9. After spending a third of my money I find I have 3*s.* 4*d.* left. How much had I at first?

10. Two girls had tea at a tea-shop. The waitress charged for both on one bill, which came to 2*s.* 6*d.* One girl had a threepenny cake more than the other. How much of the half-crown ought each to pay?

11. A certain man can dig a garden in 3 days, and his son can do it in 6 days. If they both work together how long will they take?

12. Nine soldiers eat their food in one hut and 15 in another. Seven loaves of bread are allowed for each hut. If the 9 soldiers in the first hut eat their bread in 4 days, how long will it take the 15 soldiers to eat theirs, assuming that all soldiers eat the same amount?

13. A bankrupt pays 5*s.* in the £. What did he owe to a creditor to whom he pays £16 10*s*?

14. After the population of a town has decreased by 10 per cent. the number of people left is 18,000. How many were there at first?

15. A shopkeeper wishes to sell a blend of tea at 2*s.* 3*d.* a pound by mixing tea at 2*s.* a pound with tea at 3*s.* a pound. In what proportion should he mix them?

16. If I lose 5 per cent. by selling an article for £9 10*s.*, what should I lose or gain per cent. by selling it for £10 5*s.*?

17. A, B and C cycle continuously round a circular track. A takes 8 minutes to go round, B 9 minutes, and C 12 minutes. If they all start together, how long will it be before they are all together again at the starting-point?

18. P and Q are two towns 210 miles apart. A train which travels at the rate of 30 miles an hour starts at 9 a.m. to go from P to Q. At the same time a train which travels at 40 miles an hour starts going from Q to P. When will these trains meet?

19. A man rows at the rate of 5 miles an hour with the stream, and 2 miles an hour against the stream. What is the rate of the stream?

20. A cyclist who rides at the rate of 12 miles an hour chases a man who walks at the rate of 3 miles an hour and who has had 2 hours' start. How long will it take the cyclist to catch the walker ?

The norms are as follows—

Age.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.
Score . .	2	$3\frac{1}{2}$	5	$6\frac{1}{2}$	8	$9\frac{1}{2}$

The lowness of the level of achievement will probably surprise those who have not had a wide experience in setting arithmetical problems to children. It is possible that the results I obtained were somewhat vitiated by war conditions, and that it is desirable to revise the norms later on when the schools have completely recovered. Indeed, it is always expedient to make a periodical revision of scholastic standards, for these standards are the resultant of native intelligence, courses of study, and efficiency of teaching ; and two at least of these factors vary with time.

The boys do better than the girls in applied arithmetic, but the difference in these simple problems is too slight to justify separate norms for the two sexes.

CHAPTER XI

PRACTICAL ABILITY

We cannot test the mind as a whole : we can only test it piecemeal. And in testing it piecemeal, it is found that certain specific abilities manifest a sort of kinship. They not only seem on the surface to be kindred abilities, but when tested they give results which are highly correlated. They form a natural group, clearly separated from other natural groups. Of the three such groups which are most readily discernible, literary ability seems to be the factor common to one, mathematical ability to another, and motor ability to the third. It is this third factor, the factor that underlies success in drawing, handwriting and handwork, that here demands our attention.

It seems reasonable to assume that success in all sorts of manual work depends on the general efficiency of one's motor apparatus. A man's body is many things at once. To the food-faddist it is a digestive tube, to the bacteriologist it is a happy hunting-ground for bacilli, to the boxer a fighting machine, to the philosopher a thinking machine, and to the craftsman a wonderfully complex and delicate instrument for making things. And the important point of application of the instrument is the hand. The hand is in fact taken as representing the whole

neuro-muscular system. And if we test the dexterity of the hand we are supposed to test motor ability as a whole. We must distinguish, however, between innate ability and acquired ability; for it is the former that mainly interests the psychologist. The primary and basal test of motor ability has for its aim the discovery, not of the amount of motor skill that has been acquired by practice, but of the amount of aptitude for acquiring skill. It is the potentiality we wish to measure, not the actuality; the original capital, not the interest that has accumulated through careful investment. Just as the aim of Binet's tests is to measure that general intellectual ability which is independent of schooling, so the aim of a motor test is to measure that general motor ability that is independent of training. And the only test of this kind that has crept into general use among psychologists is the tapping test.

Essentially the test consists in discovering the number of taps the unpractised subject can make per second. A pencil and a piece of paper roughly serves this purpose. It has, however, two disadvantages: the marks made are difficult to count, and the subject cannot be allowed to tap more than once on the same spot. For careful work a piece of apparatus is necessary which records automatically the number of taps made, whether made with a stylus on a flat surface, or by merely pressing a small lever after the manner of the telegraph operator. The claim of tapping to be regarded as a test of general motor ability is considerably strengthened if the results resemble in a broad way the results of intelligence tests—if they show a gradual improvement with age, if the maximum is reached during adolescence,

and if the effect of special training is negligible. The first two conditions are satisfied in the outcome of a research conducted by Miss Bickersteth at Oxford, and recorded in Vol. IX of the *British Journal of Psychology*. She tested girls from five to fifteen years of age and found that as they got older they gradually improved, from 3 taps per second at five years of age, to 5 taps per second at fourteen. Here the maximum was reached: girls of fifteen did no better than girls of fourteen. The third condition, that the score should be unaffected by practice, is fairly well satisfied. Miss Bickersteth found that the improvement produced by practice was very slight. Unfortunately, however, there is a marked disparity among the results obtained by different investigators. This is largely due to the fact that they use different instruments. The instrument I have myself used gives a higher score than that used by Miss Bickersteth. My score at the first trial was 7·4 taps per second. With a pencil, however, on a piece of paper my rate is 5·4. Several adults whom I have tested with the lever instrument score about 6·5, and girls of fifteen give on an average the same score as adults. It is a curious fact that the few expert pianists whom I tested did not tap any faster than the average. Neither did expert typists.

Let us now consider a different kind of motor test, which is carried out with a piece of apparatus invented by Mr. McDougall of Oxford. The apparatus consists of a heavy brass plate with 24 raised sockets 2 centimetres high arranged in a circle. The subject has to insert a small steel plunger with a wooden handle into each socket as fast as he can;

and the time taken to go completely round gives the score. As the plunger exactly fits the socket, the skill required is the same as that involved in putting a key into a lock. Miss Bickersteth, who has experimented with this instrument, found that children of five took on an average 28 seconds, and children of twelve took 20 seconds. Beyond that age there was no improvement. Here again we have results following the same laws as those obtained by tapping.

When, however, we compare the tapping scores with the plunger scores, we come against the amazing fact that there is virtually no connection between the two series. What slight correlation there is gradually decreases with age. In other words, a child who does well at the tapping test is just as likely to do ill at the plunger test as to do well. So that if one of these tests is a test of motor ability, the other is not; for they give contradictory results. Looking closely at the two tests, however, we see that the plunger test introduces a new element. Tapping measures speed of movement only: the plunger test measures precision as well. In tapping we may aim as carelessly as we like, but with the plunger we must aim precisely. Moreover, tapping can be done blindfold, but the plunger test needs the co-ordination and co-operation of eye and hand. Motor ability, in fact, in any profitable sense of the term, is not a simple thing: it comprises at least three elements—strength, speed and precision. Anybody who goes to a fair can roughly test his strength at a machine and his accuracy of aim at a shooting booth; but if he wants it done scientifically he must go to a psychological laboratory. There he

will find a dynamometer and an ergogograph for testing his strength, and a dotting machine for testing his precision.

Another simple motor test similar to the tapping test consists in the dealing of cards. The score is given by the number of cards dealt in a fixed time. Mr. Burt has used this test extensively, and has shown that it is positively correlated with other motor tests. It is found, too, as any one who has watched an old whist player dealing cards will readily believe, that practice affects the rapidity with which it is done.

There is one important point upon which all those who have carefully investigated these simple motor processes are agreed; and that is that the connection between the simpler forms of motor ability and intelligence is much smaller than that shown by more complex motor processes. It is true that certain American investigators have arrived at different conclusions—have found a marked direct correspondence between mental and motor efficiency. But then they obtained their data from non-experimental sources. The only American who has investigated this matter by rigorous experiment is Bagley, and he found an inverse correspondence. The most careful researches of all have been made in England by Mr. Burt, Mr. Moore, Dr. Carey and Miss Bickersteth, and they all found a comparatively small amount of positive correlation between the two types of ability—an amount which got smaller and smaller as the subjects grew older, and more intelligent. With the older defective children the correlation is still high.

This is not very surprising when we discover that

the kind of ability revealed by these instruments—the tapping apparatus in particular—has apparently little connection even with skill in craftsmanship. From the few experiments that I have made (they are far too few to be conclusive) the children who are regarded as "clever with their fingers" cannot as a rule tap more rapidly than those whose "fingers are all thumbs." The only inference we can legitimately draw from these facts is that the simple motor tests I have described measure some specific ability which is not an index either of fine powers of craftsmanship or of high intellectuality. We certainly have no right to infer that the ability in question is worthless. For the same line of reasoning would force us to admit that memory too was worthless. The correlation between rote memory and intelligence is not high: a fool sometimes has a good memory, and a genius a bad one. Yet we are bound to concede that the fool would have been a bigger fool if his memory had been worse, and the genius would have been a greater genius if his memory had been better. So with natural motor capacity, in however narrow a field it may work. There may be other things that are more serviceable even for technical skill; but in itself it is good: it is better to have it than not to have it.

The only use of tapping that cannot be impugned is as a means of determining whether a pupil is congenitally right-handed or left-handed. I myself am an inveterate dextral: I tap 7·4 to the second with the right hand, and 5·5 with the left.

In experimenting with this instrument one cannot fail to be struck by the evenness of the scores. I dealt with five adults in succession who made

precisely the same score—200 in half a minute. In Miss Bickersteth's experiments the mean variation for both the tapping and the plunger scores was surprisingly low as compared with the variation in the other abilities tested. In fact, we here seem to be testing a gift which among civilised races nature has distributed with rare impartiality. People differ very little in their basal motor endowment: it is in the use to which they put this endowment in the interests of the higher intellectual powers that wide differences appear.

We cannot help feeling that we have not yet reached the root of the matter—that we have failed to put our finger on the essential thing in the making of a craftsman. And if native hand-skill is not the essential thing, what is the essential thing? Let us examine a few significant instances. Vierge, a celebrated black-and-white artist, had the misfortune to lose the use of his right arm. In a short time he was producing drawings indistinguishable in executive skill from those which he had formerly produced with his right. Mr. H. Weaver Hawkins, a young student at the Camberwell School of Arts and Crafts, was severely wounded during the war by shrapnel passing through his right arm-pit and penetrating his shoulder-blade. Pyæmia set in and attacked both shoulder-blades and both elbows. Surgical operations became necessary: the two elbow-joints were removed and artificial ones put in their place. The result was that when he got up from his bed of sickness he found his arm movements woefully restricted. He had little power in his shoulders, little control over his elbows, not much more over his wrists and fingers. He seemed like a man with

two paralysed arms. But he soon found that he could move his left hand in one series of directions, and his right hand in another series of directions. So when he returned to school eighteen months ago, and took up his art studies again under Mr. Walter Bayes, he held his pencil in his less crippled hand—his left—and by eking out its movements with his right was able to make a start in re-learning the art of drawing. Now he can draw as well as he ever could. Indeed, his work, recently exhibited at the Goupil Gallery, was highly praised by critics who knew nothing of his infirmity. Requiring two hands to do imperfectly the work of one, and drawing as it were with his whole body, he manifests a strength and precision to be found only in the accomplished artist. Here we have a notable instance of the triumph of mind over matter. To thwart the creative impulse in such a man is impossible—unless you kill him. If his hands are gone, he will draw with his feet; if his feet, too, are gone, he will draw with his elbow, his chin, his teeth—with any part of his person to which he can attach a pencil or a brush. If the essential spirit is within him it seems to create the machinery with which it works. In the long run it is brain that counts, not muscle. It is in the mind of man that artistry and craftsmanship reside: they depend on a form of psychic or cerebral energy which flows out through the hand, the hand being the most permeable outlet; but if this outlet is blocked the imprisoned energy will force open some other channel of escape.

If, therefore, we try to gauge a person's capacity for making things by getting him to wag his finger, we must not be surprised to discover that we are

merely touching the fringe of a great matter. And whatever name we may give the primary essential we may be sure it is not a simple thing. For we shall find motor elements inextricably mixed up with intellectual elements of the highest order, and emotional elements of the subtlest form. To call the central aptitude constructive ability is certainly to connect it with one of the great instincts of humanity, but at the same time it masses under one name a number of distinct abilities which we have neither analysed nor measured; and until we have measured them individually there is little hope of our being able to estimate them collectively, or, indeed, lay hold of that single factor (if there is such a factor) which may be supposed to be common to all these units. It has, however, been argued that as all construction in the handicrafts consists in adapting and arranging objects in space, a capacity to conceive and to picture those arrangements in the mind is at least part of the general constructive ability. One of Binet's intelligence tests for adults consists in folding a square of paper twice, cutting a triangular notch in one of the sides, and asking the subject to draw what it would look like if it were opened.

A more complex construction test is that of the dissected cube. The subject is asked to imagine a painted cube of three-inch side cut into cubic inches, and is required to say how many of these little cubes are painted on three sides, how many on two, how many on one, and how many on none. Another way of administering the test is to present the twenty-seven small cubes all mixed up and ask the subject to put them together so as to form one large

cube painted all over. A record is made of the plan adopted and the time taken. The subject who works in accordance with a premeditated plan is regarded as superior to the subject who works by trial and error. It is a difficult test, which can only be done with facility by what Terman calls "superior adults."

The commonest type of construction test is the form board, of which there are many varieties. The essence of the problem is the same as that of the jig-saw puzzle : how to put things together to form a whole. The time taken gives the score.

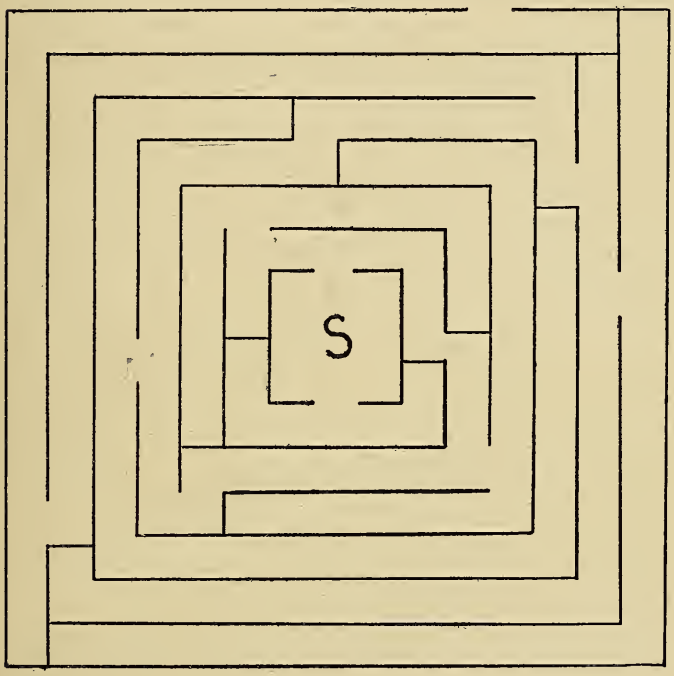
Mr. T. L. Kelly of Texas University has devised a construction ability test on the meccano principle. A quantity of material consisting of strips of wood of different shapes and sizes, and a variety of blocks and wheels, is placed before the pupil and he is asked to build the best thing he can out of them. Here initiative and inventiveness are brought into play. The test is reasonable enough and simple enough ; the difficulty lies in assessing the result—the same difficulty in fact that we have in working the models made in the handicraft centre. It is hard enough to mark them when all the things are the same : it is harder still when they are all different. In America they try to get over the difficulty by forming a standardised scale of specimens which have already been marked in accordance with the average judgment of a number of independent examiners. Any particular object to be marked has to be compared with the standard specimens until its equal is found. But it is not easy to find its equal. It is not easy to say whether an indifferently made toy table is equal to, better than, or

worse than, a badly made parquetry mat. The specimen scale, in fact, does not do away with individual judgment, nor does it appreciably lessen the variability of marking.

The test that seems to give the best evaluation of the factors which make up practical ability is the maze test. Mr. S. D. Porteus, the Director of Research at the Training School, Vineland, New Jersey, has invented a series of these tests, has standardised them, and has compared the results with those obtained by using other types of tests. The correlation coefficients show that the Porteus tests, while yielding as a whole the same type of developmental curve as Binet's, differ widely from them in the estimates they afford of certain individuals who are mentally unstable and are temperamentally unfit to make their way in the world. Mr. Porteus, in fact, contends that the Binet tests are too exclusively intellectual, and that they generally fail to detect the children who have brains but no grit, who have linguistic ability but little common sense, who show a superficial brightness but have little capacity for forethought and planning.

The accompanying diagram represents the Porteus test for children of 14 years and 6 months. It is one of the series of eleven standardised for the years from $3\frac{1}{2}$ to $14\frac{1}{2}$, no test being given for $13\frac{1}{2}$. The problem is to trace with a pencil the most direct route of threading the maze by entering at S and getting out again by some other exit. The following are Mr. Porteus's instructions: "Show child starting-point at S, and tell him to find his way through the test form without going along any blocked path. As soon as a mistake is made, stop

the child and bring him back to the starting-point for his second trial. Never allow the child to retrace his course." Half-year credit is allowed if the test is passed on the second trial. If he passes on the first



Porteus Test Year XIV

trial his mental age is regarded as at least $14\frac{1}{2}$, if on the second at least 14.

A new line of investigation has recently been opened up in America, mainly by F. B., and

L. M. Gilbreth. It is technically termed "motion-study." Mr. Gilbreth made a systematic study of brick-laying, and found that a big reduction could be made in the number of separate movements involved. From eighteen they could be reduced to five; with the result that 30 men could by the new method lay as many bricks as 100 men by the old method, and would be less fatigued at the end of the day. A brief account of this research will be found in Dr. Myers's *Present-Day Applications of Psychology*. This little book also describes the chronocyclegraph, an instrument for photographing a movement so that it can be examined in all its details, and the extent and speed of each part scientifically measured. By means of it an awkward action may be compared with a skilled action, so that the needless and harmful elements may be brought to light and eliminated. It bids fair to reveal to us in due time how we may learn to work efficiently and to play efficiently; and perhaps allows us to indulge in the Utopian hope that the distinction between work and play may disappear altogether. With the drudgery part reduced and the creative part increased work will become the joy and delight we all feel it ought to be.

The conclusion at which we arrive is that skill in carrying out any piece of practical work, needing as it does the thinking mind as well as the creating hand, involves a large number of special aptitudes and a large number of special habits; and that mere motor ability, in the barest sense of that term, is only a part, and that not the most important part, of the whole process. Attempts to measure innate dexterity have not so far proved of any help to the

teacher; nor have the methods of measuring the higher functions involved in constructive work developed sufficiently to give the young craftsman any guidance in pursuing his craft. We have not yet discovered a good indirect means of testing practical ability: we must test it directly. We must measure it as we have always measured it, by getting the pupil to do a piece of work and forming our estimate of it by standards which knowledge and experience have fixed in our minds.

CHAPTER XII

COMPOSITION

There is no branch of study more important than English Composition, and there is none so hard to mark. So heavy, indeed, is the task, that teachers have been known to bargain with their chiefs about the minimum amount of marking to be done. In London Elementary Schools the marking of one set of composition exercises per week is generally regarded as all that can be reasonably demanded of the class master. Some teachers do more, but this is looked upon as a work of supererogation. Marking involves a twofold task—correcting the imperfections and appraising the result; and of these, one seems endless and the other hopeless. And the corrections do not help us much in aiming at an estimate of the merit of the exercise. For to appraise a piece of writing by counting the blunders is itself a blunder. On this basis Shakespeare would fare badly, and Lamb would come off worse than if he had placed himself in the hands of the schoolmaster who wanted to teach him how to write essays. A pretentious piece of writing may have nothing in it—not even blunders; and another piece may be full of good things, and at the same time full of faults. Not that faults do not matter, but that value is to be judged

positively, not negatively—not by subtracting marks for each fault, but by adding marks for each merit.

And the worst of the business is that the marker, after spending weary hours in toiling through a pile of papers, cannot rid himself of the suspicion that his labours are vain. He knows that the pupil, ignoring the emendations in red ink, will simply glance at the final mark and cast the paper aside. The corrections are made in the wrong place. The mistakes are removed from the paper, whereas they should be removed from the pupil's mind. The corrections are not made in the right place, because they are not made by the right person. For the right person is the pupil himself.

How, then, can we make the task of correcting fall more upon the shoulders of the pupil who is benefited by it, and less upon the shoulders of the teacher who is bored by it? To begin with, the pupil must have ideas. He should be put to write on those topics only which really occupy his mind. Every child thinks about something—constantly thinks about something—and that something, if it can be discovered, is the theme upon which he is best fitted to write. His ideas are best put when they come warm from his brain. Then only does writing become a means of self-expression. But since his ideas are often meagre and trivial, and the composition exercise is a means of enlarging both his circle of thought and his store of words, as well as of improving his control of what words he already has, he should constantly be encouraged to venture on new topics. And for this an opportunity for preparation is essential.

Then, again, when he has finished the exercise he should be given a chance to revise it. Many of the mistakes which we point out to him he could quite well with a little trouble find out for himself. When you or I write anything, rarely do we leave it as it first flowed from the pen. We set it aside and read it again later; we score out superfluous words, change awkward phrases, rearrange the ideas, and sometimes, indeed, write the whole thing over and over again. All careful writers do this. If they do not actually do it on paper, they do it in their heads before committing it to paper. To revise and to remodel, to reflect upon what is written, and to reject even the good in favour of the better—that is at least part of the secret of clear and vigorous prose. We do this ourselves, but do not allow our pupils to do it. Often do we expect them to write without preparation; always do we expect them to write without revision. Second thoughts are discouraged; for erasures are discouraged. The pupil must try to present a fair page of writing without blot or blemish. So chary are the children of crossing anything out that if they make a mistake in phraseology or in spelling, they enclose the peccant word in brackets and leave it there. Let the teacher countenance, nay praise, the untidy page (provided the untidiness is due to careful thinking, not to careless writing), and he will find his pupils falling into a habit of self-criticism. If the writing has become illegible the piece should, of course, be re-written. The stages of a composition exercise are, therefore, three: preparation, rough draft and final copy. And they may require three distinct lessons, or

any two sequent stages may occupy one lesson, or any one stage may spread over many lessons.

So much for correcting. The marking proper, the measuring of the achievement, still remains to be done. And in that lies the crux of the difficulty. So complex is the thing to be measured, so numerous the criteria of merit, so diverse the points of view, that it is almost impossible to find different examiners of the same scripts arriving at the same marks. They form different estimates because they attach different values to the component factors. Some think highly of quantity, others of quality; some of ideas, others of style; some of wealth of words, others of clearness of diction; some of logical arrangement, others of sound and rhythm. Indeed, there is no end to the qualities which may be exalted by some and belittled by others. In measuring the achievement we are measuring a medley. It is as though we were trying to represent the value of a room by a number which should sum up the size of the room, its shape, the lighting, the ventilation, the convenience of the fittings, the pattern of the wall-paper, the state of the floor, and a host of other qualities and quantities. The consequence is that in appraising a piece of writing we rely on the general impression left on our minds after reading it. In fine, we do not measure at all: we guess. True, it is not a pure guess: but neither is it a pure guess when we guess the height of a building by looking at it. We have data to go upon, but our mode of estimating is purely subjective. The standard by which we judge is our own standard and nobody else's. And the aim of the modern movement of reform in testing is to

provide objective standards—standards that are everybody's standards.

In America they try to objectify the standard in composition by means of a scale of specimens. Certain pieces are standardised by a number of reputable examiners; they are arranged in order of merit, and assigned definite marks. This is then used as a scale by reference to which composition exercises are measured. Five such scales are in general use, of which the Hillegas-Thorndike is apparently the most popular. This scale has 15 grades of merit, ranging in marks from 0 to 95—presumably out of a maximum of 100. The sixth specimen from the bottom reads as follows—

"De Quincy.

"First: De Quincy's mother was a beautiful women and through her De Quincy inhereted much of his genius.

"His running away from school enfluenced him much as he roamed through the woods, valleys and his mind became very meditative.

"The greatest enfluence of De Quincy's life was the opium habit. If it was not for this habit it is doubtful whether we would now be reading his writings.

"His companions during his college course and even before that time were great enfluences. The surroundings of De Quincy were enfluences. Not only De Quincy's habit of opium but other habits which were peculiar to his life.

"His marriage to the woman which he did not especially care for.

"The many well educated and noteworthy friends of De Quincy."

This specimen is labelled Quality 47. What are we to think of it? It reads as though the writer had been put to read something about De Quincey, had ill-digested it, and had tried to reproduce it; not because he wanted to but because he had to. It may be typically American; it certainly is not typically English. As it stands half-way up the scale we are, I presume, to regard it as an average performance. But the average performance in our English schools bears so slight a resemblance to this specimen that the two are in no way comparable. The bulk of the children's writings strike a more even and a more genuine note. We never find a pupil using such booky words and showing at the same time so feeble a grasp of the structure of the sentence. If he can use the colon (he rarely can, by the way) he will not bungle at the apostrophe. If he uses long words he generally knows how to spell them. Indeed, it would be difficult to pick up in our schools a piece of composition which, exhibiting the same combination of merits and defects, could be confidently judged as equal to the standard sample, and therefore deserving of forty-seven marks out of a hundred.

There is the further difficulty of comparing disparate types of composition. To equate an essay with a story is by no means easy. A realisation of this difficulty led to the formulation of the Harvard-Newton scale, which comprises four distinct series of samples, one for each of the four common forms of discourse: narration, description, argumentation and exposition. But the whole question of the value of a scale of samples is still open. It is certain that its worth has never been clearly demonstrated,

In one group of investigations upon one particular scale, indeed, it was actually found that teachers who used a scale of this kind showed greater variability in their marking than those who adopted the usual plan of relying on general impression. In fact, they measured better without the scale than with it. On the other hand, it is claimed that with practice this variability decreases, so that after a time the advantage in steadiness will lie on the side of the scale. Be that as it may, its merits are not so obvious as to lead to its adoption in other countries. If it ever comes to England it must put on English garb. The sample pieces must be taken from English schools and standardised afresh.

A promising line of inquiry has been opened up by Dr. Kimmins in his inquiry into the methods of expression used by London children in essay writing at different ages. (The *Journal of Experimental Pedagogy*, III, 289.) The criterion used by Dr. Kimmins is the type of sentence employed. From this point of view the most significant change with increasing age seems to be the less frequent use of the simple unrelated sentence, and the more frequent use of the complex sentence. But important as is this way of looking at children's writings, it is not the only way: it merely marks one point of merit out of many.

There is, in fact, no help for it: we must, for the present at least, fall back upon the method of personal impression. And, indeed, when we remember that in addition to the more tangible features of a piece of writing, there is always that peculiar appeal to our æsthetic sense which defies all measurements and all standards—when, in fact,

we remember that it is an artistic product as well as an intellectual product—we find it hard to see how the subjective standpoint can ever be outgrown. There are, however, certain definite things we can do to bring our individual judgments into closer unison.

First, we must realise the roughness of the scale we are capable of using, and give up trying to calibrate more finely than the conditions warrant. To pretend to find 100 degrees of difference in however huge a number of papers, is to expose oneself to ridicule. Who can say, with any measure of confidence, that one paper merits 58 marks exactly and another 59?—unless, indeed, he works on some mechanical system of adding sentences, of deducting for errors, and of ignoring all the broader and more spiritual issues. The most finely graded scale in America, the Hillegas-Thorndike, has only 15 degrees of merit; the others have respectively 10, 9, 8 and 6. At our universities it has ever been the custom to grade essays in four groups, assigning to each one of the first four letters of the Greek alphabet, with an occasional plus or minus to indicate finer shades; and even this broad classification inspires in the undergraduate no great degree of confidence, maintaining, as he often does, that—

> "'Twixt right and wrong the difference is dim,
> 'Tis settled by the moderator's whim;
> Perchance the delta on your paper marked
> Means that his lunch has disagreed with him."

Having fixed upon the number of grades of merit (five is probably enough to start with) we must

criticise the efficiency of our marking by observing the way in which our scores are distributed. If, for instance, we use five grades the coefficients in the expansion of $(x + 1)^4$, that is, 1, 4, 6, 4 and 1 give the probable or normal distribution. Generally speaking, out of every 16 papers 1 should receive one mark, 4 two, 6 three, 4 four and 1 five. If ten grades are used, the expansion of $(1 + x)^9$, that is, 1, 9, 36, 84, 126, 126, 84, 36, 9 and 1, apportions the number of papers that should, in a reasonable system of marking, receive the scores from 1 to 10 consecutively. This, however, is a guiding principle, not a compelling principle. It helps the teacher in cases of doubt, it shows up the defects of a faulty system, but it cannot decide what marks are actually merited. The real starting-point is neither the best paper nor the worst, but the middle or average. The marks should crowd round this middle, but at each end there should be elbow-room. This average is the standard which the marker should have fixed in his mind, and with which he should mentally compare the individual papers. There must be much provisional marking before this standard is fairly established. And in the case of the class teacher it is a standard which should be constantly rising, and should be assigned a constantly increasing value.

Having developed for himself a sensible system of marking, the teacher should now impart it to his pupils. He should teach them how to mark. It will save him hours of useless toil, will alter the pupils' attitude towards his judgments, and will evoke in them a healthy spirit of self-criticism. At present the schoolboy never thinks of challenging

the teacher's verdict. He regards it not as a valuation but as a gift. He asks himself, "What mark has he given me?" He should learn to ask, "What mark have I earned?" And he should be able to say approximately whether his wage is correct, and have the right to challenge the figure.

One of the pupils should occasionally read his composition to the class, and the rest be required to record individually and independently their estimate of its merit. After a little practice it is surprising how close together the estimates become. The scholars tend more and more to agree with one another and to agree with the teacher; and soon the marking of a complete set of papers may safely be delegated to one of the scholars. He may be trusted to take pains over the task, knowing as he does that his marks will be carefully scrutinised by the writers, that his corrections and his findings will in some cases at least be challenged, and that he is to-day judging the work of one who to-morrow may become his judge. When the marked papers are distributed there will probably be much commotion in the class. There will be a simmering of indignation and protest. And if the teacher has "nerves" he had better not try the system. With patience and tact, however, all the troubles will disappear. The teacher becomes the umpire between the plaintiff, who states his grievance, and the marker, who has to defend both his corrections and his assessment. And the discussion that arises will be of more value to both pupils than much red ink and much blue pencil. In course of time the class will become educated to this sort of thing, and the members will take kindly and calmly the

criticisms of their fellows. It affords the same sort of moral training as boxing: it teaches them to take hard knocks without losing their tempers. Moreover, they are really learning to write English.

No claim of novelty or of originality is made for this system. In modified form it has been used by others, and used with signal success. Nor must the teacher think that it entirely relieves him of the burden of marking: it only reduces the burden. In all cases he is the final arbiter: in important examinations he is the sole arbiter. And when he does mark he should mark very carefully; for his papers will now go back to trained critics.

APPENDIX

SOCRATES ON INTELLIGENCE

THERE is little doubt that the dream itself was due to the lunch; but the content of the dream was determined by other things. It was holiday time. In the morning I had taken a long and solitary walk; and, pondering over the question of the discipline of the mind, had followed trains of thought which led to flatly contradictory issues. After a somewhat hearty lunch I retired to my study and took down from the shelves a volume of Jowett's translation of Plato. The book rested on the broad arm of my reading-chair, and as I turned over the leaves I was overcome with drowsiness and fell into a profound sleep. And as I slept I dreamed a dream. And I thought I stood in the streets of a strange city, which by some obscure process of reasoning or intuition I knew to be the ancient city of Athens. And of those who passed along the sunlit street two men specially arrested my attention. One of them was short and ungainly, with a snub nose and protuberant eyeballs. His feet were bare and his simple cloak old and weather-stained. Indeed, his unattractive aspect formed a marked contrast with that of his companion, who, although a somewhat older man, suggested by his dress and general bearing the

old Greek ideal of a gentleman—*καλοκἀγαθός*. Yet was the ill-favoured one not devoid of a certain native dignity; and I had no difficulty in recognising him as Socrates. Who the other was I could merely surmise. I followed them for some distance until they turned into a porch and knocked at a door of finely chased bronze. Presently the door was opened by a slave, and passing along a narrow hall we abruptly entered a room equipped with book-shelves, a pedestal desk and oak furniture of modern design. A severe-looking person in spectacles who was sitting at the desk rose to greet his visitors and bade them be seated.

The anachronisms of the dream are glaring and palpable; but during the dream itself they not only failed to astonish me: they entirely escaped my notice. It seemed quite right and fitting that Socrates should be conversing in English with a black-coated gentleman who quoted Tennyson. There appeared no historical inconsistency in the reference to prospectuses, school examinations and medical inspection. Nor did the fact in the least surprise me, that although I was present the whole time nobody seemed to take the slightest notice of me. Of the conversation that took place in that strange-familiar room my recollection is clear and vivid, and a faithful record thereof is herewith given—

Soc. Hearing of your fame as a schoolmaster, Sophisticus, I have come to consult you about my two sons, one of whom is nine and the other ten years of age. I have brought with me my old friend Crito, who also takes much

interest in the lads. We wish to know how best they may be trained in wisdom and virtue.

Soph. Well, Socrates, I do not think you could do better than send them to my school.

Soc. Many of my friends say the same thing; and that, indeed, is why I came to you. You will, I am sure, put it down to pardonable anxiety on the part of a father if I seek, by asking you questions, to assure myself that my sons will get at your school a sound education.

Soph. You may ask me as many questions as you like, Socrates, and I will do my best to answer you.

Soc. Tell me, then, Sophisticus, what will my boys learn in your school?

Soph. They will learn everything that an educated man should know. There is my prospectus. You will see therein the full list of subjects.

Soc. But I see nothing here about the teaching of virtue. Is not that one of the subjects?

Soph. We do not put that down, Socrates; but we do teach our pupils to be good. We explain to them the sacred writings, and we look very carefully after their morals.

Soc. And I see nothing here about wisdom. You will admit that a boy may know a large number of things and yet not be truly wise.

Soph. I readily admit that. Although it is not put down in my prospectus, that is really the supreme aim and purpose of my school. We have discarded the old-fashioned word "wisdom," and use the word "intelligence" instead; but it means the same thing. My boys get a good training; their intelligence is

awakened. That is in fact the chief way in which my school differs from most other schools. It is not a place where "knowledge comes but wisdom lingers." Our aim is not merely to prepare boys for examinations; what we really pride ourselves on doing is in producing general intelligence.

Soc. I begin to understand. But it is a pity you do not put this down in your prospectus, Sophisticus, for I have for years been looking for a school where they train intelligence, as you call it, and have never found one. I see, however, that you advertise the fact that pupils are prepared for certain examinations.

Soph. If I did not put that down, Socrates, I fear I should get no pupils at all. But what I really try to cultivate is general intelligence.

Crito. Don't you see, Socrates, that that is the explanation he reserves for the parents of children who fail at the examinations.

Soc. That suspicion, Crito, is unworthy of you, and does injustice to a great schoolmaster like Sophisticus. I indeed prefer to believe that he seeks to produce general intelligence, and that success at examinations is a sort of by-product.

Soph. You state the case truly, Socrates.

Soc. But tell me what you mean by giving the boys a good training.

Soph. I mean that we cultivate their mental powers. We teach them not only to know, but to do. We make them remember better, observe better, and reason better, and, to put it generally, we make them better thinkers.

Soc. That, indeed, is a great achievement. But you must not bewilder me by telling me so much at once, for I have a bad memory, and can only deal with one thing at a time. Tell me how you make a boy more observant.

Soph. By giving him practice in observation, of course.

Soc. By practice, you mean that he repeats the same act over and over again?

Soph. Just so.

Soc. And he does it better the second time than the first?

Soph. Yes.

Soc. And the third time better than the second?

Soph. Precisely.

Soc. But is not this what we call forming a habit?

Soph. I suppose it is.

Soc. Observation, then, is a habit?

Soph. I have never thought of it in that light, Socrates, but I think you are right.

Soc. I do not say that I am right: I am trying to find out what you can tell me about the matter. You do not mind my putting these questions?

Soph. Of course not, Socrates. I have heard of your custom of questioning people, and I am glad to have an opportunity of hearing you.

Soc. Tell me, then, does training consist in anything else but the formation of habits?

Soph. It depends on what you mean by habits.

Soc. Does not an act tend to become easier by repetition?

Soph. Yes.

Soc. And it is true of an act of thought, of feeling, or of will, as well as of a physical act?

Soph. Certainly.

Soc. Shall we agree to call any act that has been improved by practice a habit?

Soph. It seems to be a suitable name for it.

Soc. Then training consists in forming habits?

Soph. Yes.

Soc. And nothing else?

Soph. And nothing else.

Soc. And is intelligence a habit?

Soph. Well, I should not be disposed to call intelligence a habit. It seems to me to be less a habit than observation even.

Soc. And yet you said that you could train intelligence.

Soph. There does seem to be a contradiction somewhere. And yet I am sure I can train intelligence.

Soc. Let us try again. What is intelligence?

Soph. Now you have asked me a very difficult question. I can pick you out my most intelligent pupils, but I cannot tell you off-hand precisely in what this intelligence consists.

Soc. Are they necessarily intelligent if they can read well?

Soph. No.

Soc. Or write, or draw, or sing well?

Soph. No.

Soc. Or repeat poetry, or remember history?

Soph. No.

Soc. Or do arithmetic?

Soph. It depends upon the kind of arithmetic.

Soc. What kind of arithmetic can an unintelligent boy do?

Soph. Simple straightforward sums, such as the common rules which he has been taught.

Soc. And what kind can an intelligent boy alone do ?

Soph. Problems.

Soc. And what is the essential difference between the two ?

Soph. The unintelligent boy can do sums which he has previously practised, or has at least been shown how to do. The intelligent boy can do sums which are not quite the same as any others which he has done or been shown how to do.

Soc. And is intelligence shown in any other subject besides arithmetic ?

Soph. Certainly, a boy may show intelligence in geography, or history, or science, or handwork, or indeed in any subject which is not purely mechanical.

Soc. And is it the novel part that requires intelligence in other subjects as in arithmetic ?

Soph. That is right, Socrates, you have made it quite clear. It is the new, the unfamiliar part of any given situation that calls for intelligence on the part of a boy.

Soc. I am glad to hear you say that, Sophisticus, for that simplifies the case. Tell me, does training have to do with the familiar, or with the unfamiliar ? With the old or with the new ?

Soph. I do not see what you mean.

Soc. Does not training involve doing the same thing over again ?

Soph. Yes, we agreed that it was based on practice.

Soc. And when you are trained to deal with any kind of situation that situation is no longer new ? it is no longer unfamiliar ?

Soph. Quite so.

Soc. Then a trained faculty has to do with familiar material?

Soph. Yes.

Soc. And intelligence deals with the unfamiliar?

Soph. Yes.

Soc. Then intelligence cannot be trained?

Soph. No, Socrates, I will never admit that. There must be some flaw in the argument. I think I see what it is. The situation which you call new or unfamiliar is never wholly new. There is always much that is old, and rarely more than a little that is new.

Soc. But the unintelligent pupil can deal with the old part of the situation if he has been trained to it?

Soph. Yes.

Soc. And if the new part is not dealt with, intelligence is not shown?

Soph. Quite so.

Soc. Then it is only intelligence that can deal with the new part?

Soph. Yes.

Soc. And no training can enable us to deal with the new?

Soph. It seems so.

Soc. Then we are again back in the same position. Intelligence cannot be trained.

Soph. Although I cannot refute your argument, Socrates, in your own way, I can do it in another way. If you will send your boys to me you will find that at the end of a year they will be much more intelligent than when they came; and I call that a much stronger argument than even you can devise.

Soc. If you can do that I will believe that our discussion has somehow strayed from the truth. I will think over what you say. Come, Crito, let us depart, for I see that Sophisticus is impatient of all this talk about things which he can do but cannot explain.

(After the usual greetings they take their leave.)

Crito. Tell me, Socrates, do you really think that intelligence cannot be cultivated?

Soc. I should be glad to be convinced to the contrary. I fear there were many errors in our discussion with Sophisticus, but not of the nature that you imagine. You noticed that we seemed to agree to call observation a habit?

Crito. I noticed that, Socrates.

Soc. And do you think it is a habit?

Crito. I thought so at the time, but I don't feel so sure about it now.

Soc. Tell me, Crito, is breathing a habit?

Crito. We never call it a habit.

Soc. But is it not an act which we are continually repeating?

Crito. Certainly.

Soc. How, then, does it differ from an act which everybody calls a habit, such as the trick which some people have of frequently stroking the beard?

Crito. Only some people stroke their beards, but everybody breathes.

Soc. And why does everybody breathe?

Crito. He cannot help it, Socrates. It is a natural power which he brings with him when he comes into the world.

Soc. And why does not everybody who has a beard stroke it?

Crito. Because it is a habit which some acquire, and some do not.

Soc. A habit, then, is a personal acquisition?

Crito. Yes.

Soc. And in that sense breathing is not a habit, but a natural power?

Crito. Yes.

Soc. Can a natural power be improved, do you think?

Crito. I think it can, Socrates. My grandchildren who are now at school have breathing exercises every day. And the physician visits the school, and examines the children's noses and air passages, so that obstructions may be removed. People, too, who have their voices trained say that attention is paid not to their throats where the sound is produced, but to the mechanism for breathing. They are in fact taught to breathe better.

Soc. Tell me any way in which children at school are taught to breathe better.

Crito. They are taught to use handkerchiefs, and to breathe through the nose, and to breathe deeply.

Soc. And is not the proper use of a handkerchief a habit?

Crito. Certainly.

Soc. And breathing through the nose, rather than the mouth, is that a habit, too?

Crito. It is a habit.

Soc. And what shall we say for breathing deeply? cannot that become a habit?

Crito. It can.

Soc. Then, shall we be right in saying that a natural power like breathing can be improved by forming certain habits which render that power more effective?

Crito. We shall be right in saying that.

Soc. And is not the same thing true of observation, which we have agreed to regard as a natural power?

Crito. Yes.

Soc. And is not intelligence a natural power inasmuch as everybody possesses it in some degree?

Crito. Yes.

Soc. Then intelligence can be improved in the same way as breathing and observation can be improved?

Crito. It seems so.

Soc. Then Sophisticus is right after all, and intelligence can be trained.

Crito. And yet, Socrates, it seemed to me while you were arguing with Sophisticus that the opposite was true.

Soc. May not both conclusions be true?

Crito. How is that possible, Socrates?

Soc. I fear, Crito, that we have been confusing ourselves and obscuring the argument by using the same words in different senses. When we talked about observation we at one time assumed it to be a simple power which could either be trained or not be trained; at another time we spoke of it as though it were a group of distinct and separate powers, some, if not all, of which could be trained. And in the same way we have been deceived by different

meanings of the word intelligence. I think, Crito, that you and I had better go to school again to learn the right use of words.

Crito. I am quite willing to go with you, Socrates, provided you can find somebody to teach us.

Meanwhile we were walking along narrow streets whose bare and windowless walls were relieved here and there by doors that opened outwards into the street. One of these doors, more pretentious than the rest, was protected by a porch, near the entrance of which stood one of those images of Hermes which Alcibiades was supposed to have mutilated. As we passed this porch, which we did just as Crito had finished speaking, I touched the image with my hand and it fell to the ground with a loud crash. So loud indeed was it that I woke up with a start, and looking down I saw the volume of Jowett's Plato lying flat upon the floor.

INDEX

Absurdity tests, 30, 38, 39, 76, 154
Adams, J., 109
Addition, 161 ff.
Æsthesiometer, 9–10
Age-performance, 13–14
Age test, 60
Alcibiades, 232
American tests, 15, 20, 29, 40, 46, 145, 214, 215
Analogies test, 32, 33
Applied arithmetic, 191 ff.
Arithmetic, 19, 112, 160 ff., 226
Arithmetical devices, 190
Association test, 37, 80
Average, 123 ff.
Ayres, L. P., 112, 156–7

Bagley, W. C., 200
Barème d'instruction, 40, 107, 111, 134
Bayes, W., 203
Bickersteth, M. E., 198–9, 200, 202
Binet, Alfred, 11, 13 ff., 17, 20, 23, 27, 30, 35 ff., 48 ff., 107, 111, 197
Blackboarditis, 182–3
Bottle test, 34
Breathing, 229–30
British Association, 47, 160
Brown, W., 47
Browning, R., 153
Burt, Cyril, viii, 11–12, 16, 18, 27, 45, 47, 48 ff., 90 ff., 115, 127, 157–8, 161, 200

Cancellation, 34
Carey, N., 47, 200
Central tendency, 122 ff.
Coin tests, 63, 68, 72, 73
Colour test, 61
Columbia University, 15, 42
Complexes, v
Complimentary addition, 173–4
Composition, 19, 45, 87, 128–9, 210 ff.
Correlation, 17, 18, 25, 26
Counting test, 55, 62, 71
Courtis, S. A., 112, 160
Criminology, 5
Crito, 222 ff.
Cube test, 204–5

Dale system of reading, 141 ff.
Darwin, C., 5
Date test, 72
Day of week test, 63
Decomposition method of subtraction, 168 ff.
Definitions test, 64, 73, 84
De Quincey, 215
Dictation, 158–9
Dictation test, 69
Differences test, 68, 86, 88
Digits test, 36, 50, 55, 61, 73, 80
Distribution, 115 ff.
Dotting machine, 11, 46
Drawing, 132
Drawing tests, 13, 62, 75, 85, 86
Dynamometer, 7

Einstein, 2
English, H. B., 47
Equal addition method of subtraction, 168 ff.
Ergograph, 7
Error, Curve of, 120
Evans, Miss Lloyd, 153
Examinations, 2, 14, 18, 19, 27, 29, 41 ff., 110, 121 ff., 127, 131

Finger test, 62
French test, 33
Frequency-column-graph, 117

Gall, F. J., 3, 4
Galton, F., 6, 10, 17
Gilbreth, F. B. and L. M., 207–8
Goddard's revision of Binet's tests, 15
Gray, P. L., 160
Green, J. A., 47, 160

Hart, Bernard, 47
Harvard-Newton scale, 215
Hawkins, H. W., 202
Heights of Englishmen, 115 ff.
Heine, H., 5
Hillegas-Thorndike scale, 214

Images, 10
Indirect measurement, 3 ff., 26
Individual work, 142 ff., 178 ff.
Instructions tests, 40, 50, 59
Intelligence, 6,11 ff., 21 ff., 29 ff., 48 ff., 221 ff.
Intelligence quotient, 49
Interquartile range, 125 ff.

James, Wm., 27, 36
Jevons, S., 1
Jones, Emrys, vii
Judd, C. H., 112

Kansas silent reading tests, 146–7
Kelly, T. L., 205
Kimmins, C. W., 216
King John, Ballad of, 30–31
Knowledge, 22 ff., 106 ff.

Lamb, C., 4, 210
Lancashire, Reading in, 140
Lavater, J. K., 4
Lewis, E. O., 47
Lines tests, 56, 82
Lombroso, C., 5

Manual ability, 45, 196 ff.
Mathematics, 1, 17, 130
McDougall, W., 10, 11, 47, 198
McIntyre, J. L., 47
Mean deviation, 125 ff.
Measurements, v, vi, 2, 3
Median, 123 ff.
Memory, 10, 36–7, 176 ff.
Milton, J., 4
Months test, 73
Moore, R. C., 200
Morning and afternoon test, 60
Motion study, 208
Motor ability, 196 ff.
Munroe, W. S., 112
Myers, C. S., 47, 208

Name test, 51
Newton, I., 1
Normal distribution, 17, 118 ff.
Norms, 21, 108 ff., 139, 153, 160, 165, 166, 188, 191, 195
Nunn, T. P., viii, 115

Objects tests, 52, 68
Oblong test, 64
Observation, 225 ff.
Œdipus, 3, 31, 33
Ogive, 116
Oral arithmetic, 186 ff.

Pearson, K., 6, 18
Perseus, 31
Phrenology, 3–4
Physiognomy, 4–5
Pictures tests, 52, 56, 57, 66, 67, 82
Plato, 221, 232
Plunger apparatus, 198 ff.
Porteus, S. D., 206
Porteus test, 206–7
Poverty and attainments, 139–40, 189–90
Practical ability, 196 ff.
Practice in arithmetic, 179
Probable error, 126
Probability curve, 119
Problems test, 83

Quetelet, L. A. J., 17
Quartile, 125 ff.

Reaction-time, 7, 8
Reading, 70, 73, 110, 111, 134 ff.
Reading-books, 143
Reasoning tests, 16, 90 ff.

Recommendations for improving arithmetic, 185–6
Repetition test, 54, 60, 66, 69, 84
Rhyme test, 81
Riddles, 3, 29, 31, 33
Right and left test, 66
Rivers, W. H. R., 10

Samson, 3
Savoir faire tests, 70–1, 78
Scale of samples, 20–1
Scholarships, 44
Schuster, E., 47
Sensory discrimination, 9–10
Sentence tests, 74, 81
Sex differences, 140
Sex-naming test, 51
Shakespeare, W., 210
Simon, Th., 48
Skew curve, 122
Smith, M., 47
Socrates, 221 ff.
Spearman, C., 18, 24, 25, 27
Spelling, 112, 155 ff.
Sphinx, 3, 31, 33
Spurzheim, J. G., 3, 4
Standard deviation, 126 ff.
Standardisation, 19, 34 ff., 106 ff.
Stanford Revision of Binet's tests, 15
Stern, W., 49
Subtraction, 162–3, 168 ff., 187

Tables in arithmetic, 174 ff.
Tapping, 7, 197 ff.
Taylor, N., 47
Tennyson, A., 222
Terman, L. M., 15, 17, 49, 112
Thorndike, E. L., 2, 20, 24, 27, 112
Tossing coins, 119
Transcription test, 63

Vocational tests, 8

Weber-Fechner law, 8–9
Weights tests, 61, 73
Winch, W. H., 13, 18, 47, 176

Yerkes' point-scale, 15

PRINTED IN GREAT BRITAIN BY
RICHARD CLAY & SONS, LIMITED,
BRUNSWICK ST., STAMFORD ST., S.E. 1,
AND BUNGAY, SUFFOLK.

LIBRARY OF CONGRESS
0 021 338 646 3